Environmental Protection Agency

Management Information &
Data Systems Division

Data Management and Standardization Program Feasibility Study

Final Report

June 5, 1979

ARTHUR YOUNG

EPAX 8510-0013

ARTHUR YOUNG & COMPANY

1025 CONNECTICUT AVENUE, N.W.
WASHINGTON, D.C. 20036

June 5, 1979

Ms. Mary Lou Melley
MIDSD (PM-218)
Environmental Protection Agency
Room M3101
401 M Street, S.W.
Washington, D.C.

Reference: Contract No. 68-01-4640
Task Order No. 68-01-4640-18

Subject: Final Report

Dear Ms. Melley:

Arthur Young & Company is pleased to submit this final report on the Data Management and Standardization Program Feasibility Study for the EPA. This report reflects our response to the comments and suggestions received from you during the past several months. This deliverable will be followed in the near future by the Executive Summary which provides a synopsis of our recommendations for this feasibility study.

We greatly appreciate the active participation and cooperation which our project team has received from both you and other EPA personnel during this engagement.

Should you have any questions, please contact Mr. Philip Snyder or myself at (202) 828-7000.

Very truly yours,

ARTHUR YOUNG & COMPANY

By: G Mendenhall

Gerald Mendenhall
Partner

TABLE OF CONTENTS

LIST OF EXHIBITS

LIST OF EXHIBITS (cont'd)

Background & Scope

I. BACKGROUND & SCOPE

I. BACKGROUND & SCOPE

This document is the final report for the Data Management and Standardization Program Feasibility Study prepared in response to the Environmental Protection Agency Directive of Work Number: 68-01-4640-18 under Contract Number 68-01-4640.

In this chapter Arthur Young & Company describes the background behind the study initiation and resulting project scope, and the methodology employed throughout the study effort.

1. BACKGROUND

The impetus for this project came from several diverse sources and factors within EPA. These factors included:

- Desire for a common facility identifier,
- Realization that information is a resource that can and should be managed,
- Perception that EPA is in the third stage of ADP growth, and
- Need for an implementation vehicle for IRLG recommendations.

In the following paragraphs we describe these motivational factors and attempt to show how they came together for the initiation of the project.

(1) Common facility identifier

For a long period of time there has been active support from diverse areas of EPA for the development of a common facility identifier so that data collected and used in one program can also be utilized by other programs. Primary advocates for this feature were Region 3, Office of Enforcement, and Office of Water Program Operations (OWPO). Prior to this study OWPO had initiated their own project to develop a long-range ADP plan. A prominent feature of this plan is the recommendation for a common facility identifier for the Wastewater Treatment (WWT) Program.

(2) Management of Information

In both the public and private sectors there is growing realization that information is a vital resource to an organization that can and should be managed in the same manner as other resources, such as money, personnel, or inventories. Information is a tool for the management of the other resources, but there are aspects of information that must be managed in

order to assure its quality and its effective use for the total user community. These include:

- Location and responsibility for the data
- Data definitions
- Data coding schemes
- Data collection and processing.

MIDSD, as well as several program offices and Regions, felt that the information in the Agency was not being effectively managed. For information to be effectively managed the active support of top management is required. These advocates of informaton management wanted a third party to evaluate and document the status of information management in EPA in the hope that an objective opinion would help gain top management support.

(3) Stage 3 of Data Processing Growth

In the "stage hypothesis" of data processing growth advanced by Richard L. Nolan Ph.D, Stage 3 is defined as the control stage where emphasis has moved from management of computers to management of data. Some individuals perceive the Agency as moving into Stage 3 in many areas, and a study is underway to assess the stage of each major program and functional area. A key element in this stage of growth is the development and application of formalized controls. The initiation of this study, therefore, is compatible with the Stage 3 assumption.1/

(4) Interagency Regulatory Liaison Group (IRLG) Support

Douglas M. Costle, Administrator of EPA, has stated on several occasions, including the August 9, 1978, IRLG Common Codes Project Steering Committee meeting, that the Interagency Regulatory Liason Group (IRLG) effort to coordinate the regulatory and enforcement efforts dealing with chemical substances of the four primary Federal chemical regulation agencies is a top priority item. The IRLG, comprised of EPA, Consumer Product Safety Commission, Food and Drug Administration, and the Occupational Safety and Health Administration, has several projects ongoing to accomplish this coordination. One of the projects is the identification and implementatiuon of common coding schemes and

---

1/ Since the delivery of the draft version of this report in December, 1978, a study has been completed by Nolan, Nortan & Company assessing the effectiveness of EPA's automated information processing capabilities. The study was performed in relation to Dr. Nolan's stage hypothesis." The result, in part, was a conformation of the assumption that EPA was in Stage 3.

definitions for data elements vital to the sharing of chemical information. It was recognized early in the conceptual stages of the total IRLG effort that to be able to implement the coding requirements will require a coordinated, Agency-wide data management program.

These factors and forces came together in the initiation of the Data Management and Standardization Program Feasibility Study. During the initial Advisory Committee Meeting held March 17, 1978, it was decided by the Committee that the project should concentrate on two aspects or phases. The first was the definition and documentation of the need for a data management and standardization program based on the current status of data management in EPA as reflected in a specific sample of typical, major systems. The second phase would be predicated on the results of Phase I. If it was determined that there is a need for a data management program, than Phase II would address the policies, procedures and tools that would be required to implement the program. It was decided by the Committee that the project would not address standardization of specific data definitions or coding schemes since that is a major effort in itself to be undertaken on a selective basis as part of an operational data management program. The methodology used in performing the phases is presented in the following section.

2. METHODOLOGY

The Data Management and Standardization Program Feasibility Study was performed in two analytical phases:

- a requirements analysis phase to identify and determine EPA's requirements for data management and standardization; and
- program development phase to define and evalute alternative strategies for the development, organization, and implementation of a Data Management and Standardization Program in EPA.

A briefing of the findings and recommendations was presented to the EPA Project Advisory Committee and interested EPA personnel at the completion of each phase. This report is composed of the information presented at the briefings and the feedback obtained from EPA on the briefings. The methodology utilized in the performance of the project is depicted graphically in Exhibit I-1.

Phase I, the development and documentation of EPA's requirements for a Data Management and Standardization Program, was approached from both a quantitative and qualitative perspective.

(1) Quantitative Approach

The objective of the quantitative approach was to identify concrete examples of EPA's need for data management and standardization through a review of 15 representative systems. The systems were intially defined at the first advisory committee

STUDY METHODOLOGY

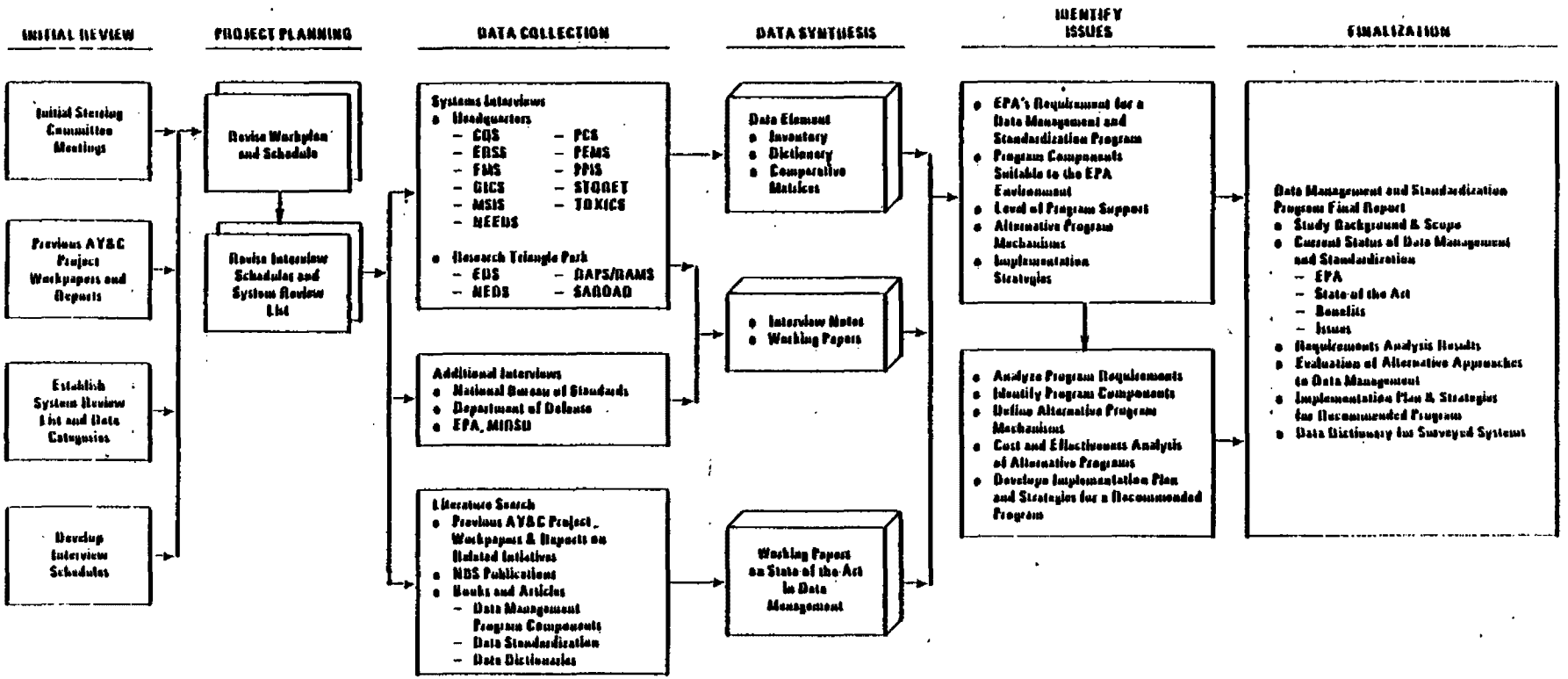


EXHIBIT I-1

meeting. Their selection was based primarily on the committees' evaluation of potential of sharing data across the systems. Other considerations included the systems visibility in EPA and that each of the Assistant Administrator's areas was represented. During the course of the project, systems were added and deleted from the sample group for such reasons as:

- Identified systems did not contain the specified data categories
- Systems were still in initial development stages
- Documentation on the systems were inadequate for complete analysis.

The systems included in the final analysis are identified both in the methodology exhibit and in Appendix A, the data dictionary prepared as a deliverable for Phase I.

At the first meeting it was also decided that for the review to be meaningful given the time and resources available for this effort, the analyses of the representative systems would be focused on five data categories. The data categories were selected by the commmittee based on their potential for being used to interface files and their current relevance to the major issues facing the Agency. The data categories selected for detailed review in the sample systems were:

- Facility Identifiers - The unique name or number used in identifying any type of facility or authority.
- Monitoring Sample Station Site - The name or identifying number of a site where monitoring or sample data is actually collected.
- Geographical Location - A location name or code which identifies the geographic location of a facility or sampling station.
- Parameter Unit Identifiers - A code on the file indicating the unit of measure for a specific field.
- Quality Assurance Codes - A code indicating the level of confidence in the data, or when the information was updated.

The systems review consisted of documentation review and supporting interviews with the appropriate EPA system's managers. The result of this process was an approximation of EPA's current level of data standardization and the need for further standardization. As a by-product, a manual Data Element Dictionary was prepared. The Data Dictionary, consisting of definitions, coding schemes, and field descriptions for the five

data categories and the sample systems, is presented in Appendix A.

(2) Qualitative Approach

The objective of the qualitative approach was to determine perceived needs for an overall EPA data management and standardization program, and potential benefits EPA would derive from the program. This approach included the following steps:

- Interviews with cognizant individuals in EPA; and interviews with external organizations such as the National Bureau of Standards and the Department of Defense, which currently maintain successful data management programs.

- A literature review to determine the state-of-the-art in information management.

As a result of this effort, the components and benefits of a data management and standardization program were defined for further evaluation in terms of their appropriateness for application in meeting EPA's requirements.

The information obtained during the quantitative and qualitative phases of the requirements analysis were synthesized and a briefing was given to the EPA Advisory Committee on July 7, 1978. This briefing was also video taped for distribution to EPA regions. A draft data dictionary was also presented prior to the requirements analysis briefing to facilitate EPA verification of the quantitative results of system review, and to stimulate concerns or ideas over the difficulties which have or could be encountered due to lack of data standardization.

The EPA Advisory Committee members were extremely helpful during the formulation stages for the approaches to requirements analysis, and also provided valuable comments and feedback related to the findings from system review, draft data dictionary, and contents of the requirements analysis. Following the briefing, direction was given to proceed with the second phase of the study.

In September, 1978, this project was temporarily suspended for one month to facilitate coordination with other EPA initiatives related to data management and standardization concepts. The initiatives included the IRLG Common Codes Project, OWPO/WWT Long-Range Plan, and Region 2 Facility File Pilot Project. Review of these other related activities further defined EPA's pressing requirements for data management and standardization, and provided insights into program implementation strategies which appeared most feasible for use from the perspective of the EPA current organizational structure.

After establishment of EPA's requirement for data management and standardization, alternative implementation strategies for a related

program were developed. These strategies were organized in terms of three program components which included:

- required organization structure for program administration and oversight,
- data management and standardization tools, and
- policies and procedures to govern program operation.

These strategies were presented to the EPA Advisory Committee on October 18, 1978, and comments were reviewed and incorporated. The effectiveness of program alternatives were assessed weighing considerations of benefits, disadvantages, impacts in the EPA environment, and implementation issues. A recommended program alternative was then selected, and estimated costs for program development, implementation and operations were developed. An implementation plan and strategies were defined for the recommended alternative program strategy.

The program development phase concluded with the delivery of the draft final report on December 4, 1978, which synthesized all study efforts to date, and presented recommendations for the implementation of a Data Management and Standardization Program in EPA. The EPA Advisory Committee decided to delay the formal release of their comments on the draft report, and thus the finalization of this report, until the results of the Nolan, Nortan & Company study, referenced earlier, were delivered. The results of the Nolan, Nortan & Company study concurred with the earlier findings and recommendations of this study and thus did not effect the Advisory Committees comments. The suggestions and recommendations made by the Advisory Committee have been incorporated into this final version of this report.

Requirements Analysis

II. REQUIREMENTS ANALYSIS

II. REQUIREMENTS ANALYSIS

In this chapter Arthur Young & Company presents the results of the requirements analysis phase of the EPA data management and standardization program feasibility study. Requirements analysis activities and findings are presented in terms of:

- Current Trends in Data Management Programs
- EPA Current Status
- EPA Requirements.

1. CURRENT TRENDS IN DATA MANAGEMENT PROGRAMS

The components of an effective data management program are presented at this point to provide a framework within which the rest of the report should be viewed. An effective data management program must consist of a dynamic program structure which actively interfaces with program participants. Our research has shown that current effective information management programs consist of three primary components:

- Organization structure which supports the program;
- Data management tools which produce the program products; and
- Policies and procedures which govern program operations.

Exhibit II-1 presents an overview of the relationship of data management program components. The functions of each of these three entities are further discussed in the paragraphs below.

(1) Organizational Structure

An effective data management program requires a defined organizational structure to implement the program, administer and monitor ongoing program operations, and initiate and implement program updates. The data management administration entity is also responsible for obtaining programmatic or user support, resolving program related issues, evaluating the ongoing benefits of program use, and seeking new areas for program application.

The actual structure for a data management program is, for the most part, dependent on the characteristics of the organization into which the program is to be implemented. Several

DATA MANAGEMENT PROGRAM COMPONENTS

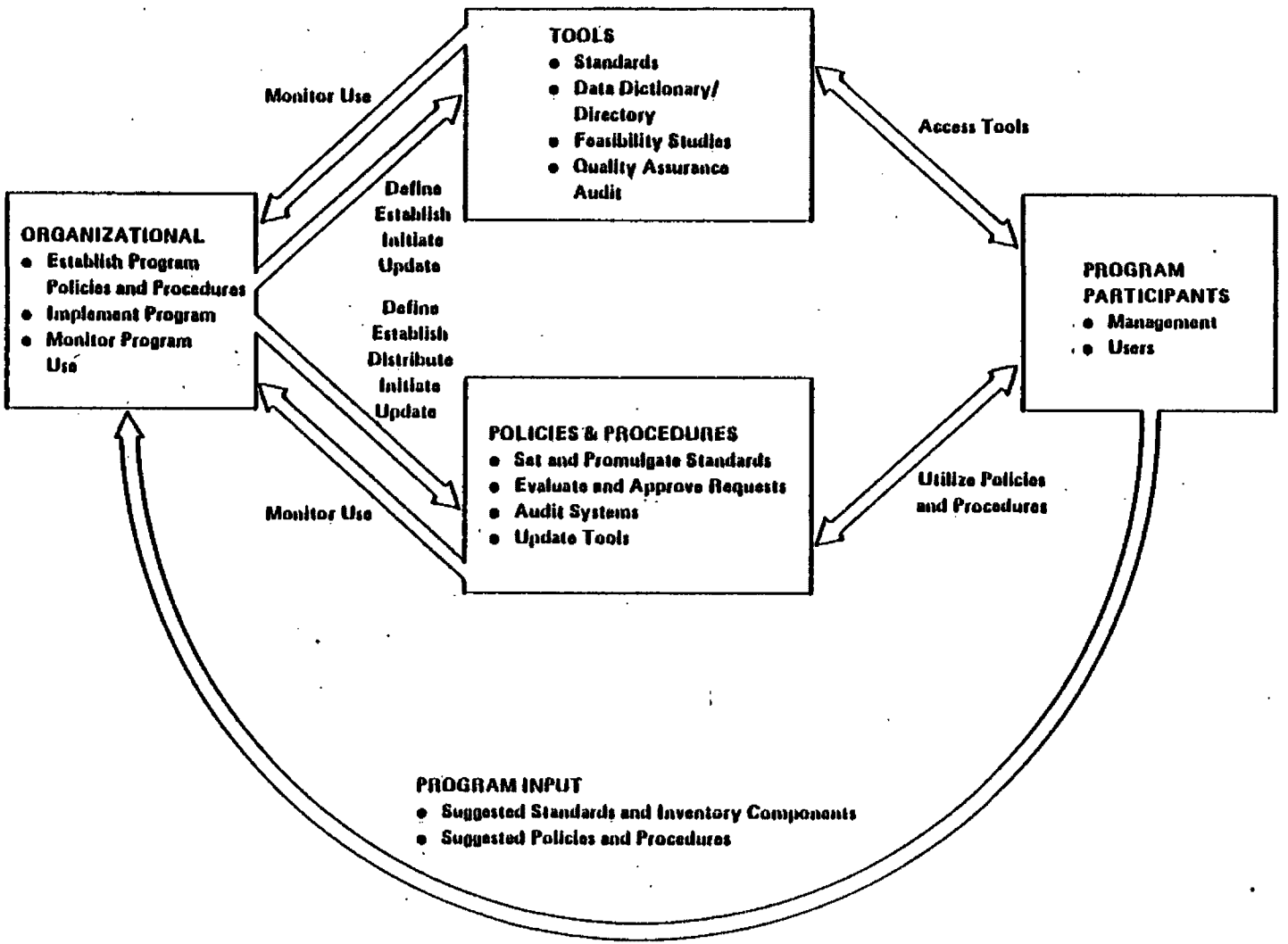


EXHIBIT II-1

factors must therefore be considered in determining what type of organizational configuration is the best strategy for successful implementation of a data management and standardization program. These factors include the following:

- Current structure of the organization including responsibilities and authorities;
- Authority levels required for successful program implementation and administration in a given organization;
- Expertise required for program implementation and operation;
- Defined program requirements;

Evaluation of these factors will lead to a determination of the proper positioning of data management and standardization program responsibilities in the organization; required functional assignments; coordination and communications mechanisms; and approval flows for program activities.

Alternative organizational structures include: a decentralized approach placing data management and standardization programs within specific units of an organization; a highly centralized program approach using a Data Administrator and staff concept which organizationally reports to senior management; or a combined or hierarchical approach. In a hierarchical approach a central policy making body, either a Data Administrator or committee, will set policies and organizational standards. The staff for this agency-wide committee will monitor and enforce the standards set by the Committee. The implementation of the program is delegated to specific units of the organization. These alternative will be discussed in greater detail later in Chapter III. Each of these alternatives must be evaluated in terms of the factors identified above.

(2) Data Management Tools

The primary tools which can be employed to effectively manage data as a resource can be categorized as follows:

- Standards
- Data Element Dictionary/Directory
- Feasibility Studies
- Quality Assurance Programs.

The use of these tools is prescribed, implemented, and maintained by the organizational component of the data management program through use of program policies and procedures.

The following paragraphs contain a more detailed explanation of these four primary data management tools.

(a) Standards

Three major types of standards can be developed to facilitate data management. These include standards for: data elements, system design and documentation, and data acquisition techniques.

- Data Element Standards - the objectives of the use of the data element standards are to:

  - Improve data accessibility
  - Facilitate timely data transfer and exchange
  - Enhance support for management decisions
  - Reduce the reporting burden
  - Improve data effectiveness
  - Increase data re-utilization.

  Data element standardization involves the use of common data definitions, data use, coding schemes, and naming conventions. For similar data elements which are found in various system files, data element standardization is necessary for the sharing and utilization of information in the files. Data transfer and exchange efforts are facilitated, and requirements for data element conversion to other forms or interpretations before utilization are reduced. This results in improved information access time and data quality. In addition, data standardization simplifies the efforts required for implementation and maintenance of a data element dictionary/directory, as well as system design and development efforts.

- System design and documentation standards - involve the use of:

  - Standardized methods for system design and documentation throughout the system development life cycle;
  - Milestones and control points for quality review of system development and documentation efforts;
  - Acceptance criteria to determine the adequacy of the system;

- System implementation procedures review to measure the feasibility of implementation strategies, user impacts, and impacts on current computer resource utilization;

- A system change approval process.

Use of these types of standards provides an essential mechanism for controlling the quality and compatibility of system development and change. System design and documentation standards result in the establishment of a detailed record of system related decisions, activities, and developments. This record can be utilized during the interim stages of system development as an information exchange medium for both system design teams and potential users, thereby maximizing use of personnel resources. The detailed documentation in a standardized format also minimizes the efforts required for system change by providing a meaningful and complete reference source for both programming and user personnel. Implementation procedures and standards reduce the risk in system development and implementation through application of proven techniques. In particular, they facilitate effective utilization of contractor support. The system change approval process assists in maintaining the integrity of both the system itself as well as any standardized data elements contained in system files.

Data acquisition techniques standards - involve the use of:

- Data collection request and approval procedures to reduce duplicate data collection

- Forms design and instruction writing quidelines to assure uniformity and clarity in data collection forms and definition.

- Document tracking procedures to provide an audit trail for locating stalled or lost source data

- Key verification techniques to control the accuracy of data entry

- Control and hash totals to assure complete and accurate processing of source data

These standards are intended to provide a control on the acquisition of data and the procedures used to handle the data prior to storage on a computer, micrographics, or other media. The objective of the control is to reduce the reporting burden on the suppliers of the source data and to improve the quality of the data received and stored. The reduction in reporting burden is accomplished in two ways. The first is by reducing or eliminating the duplicate

reporting of data. The second is by assuring that when data is requested from an entity the data definitions and coding schemes are consistent. This eliminates the need to recode information to fit a new scheme. Consistency in definitions and coding schemes will help improve the accuracy of the data reported as well. The procedures, such as the document tracking, key verification, and control and hash total will help improve quality by assuring that the data that is stored is the same as what was reported.

(b) Data Element Dictionary/Directory

A data element dictionary/directory(DED/D) is a software or firmware tool that is used to control and manage data elements in a uniform manner. It will provide a central repository of information about each data element in related systems in order to facilitate access and control of the data bases. This tool does not manage the actual content of the data, but manages the descriptive characteristics of that data (metadata). Metadata includes such physical properties as length, value range, types of admissible data, and validation criteria. Of more importance to top management is the identification of the individuals responsible for the quality and dissemination of data in the specific systems. When a request is received for specific information from other EPA Offices, Congress, GAO, or the public, management can access the dictionary to determine if the information is available, and, if it is, they can direct the query to the proper individual. Exhibit II-2 presents the basic characteristics of a DED/D and their uses.

The development of the DED/D will also provide a focal point for the development of the much needed standard data defintions and coding schemes. The DED/D will continue to be the primary tool for monitoring the adherence to the standards. This will be accomplished by reviews of new systems and elements that are being added to the DED/D.

It is possible for a DED/D to be manual using either a list or card file. Typically, because of the access and updating requirements, the dictionary is maintained via computer. The look-up media may be either hardcopy, via listings or micrographics, or through on-line query capabilities. A recent development in DED/D is the developmment of "active" DED/D systems. In an active system the DED/D dynamically interfaces with the application systems, in either a free-standing or DBMS environment, to automatically and simultaneously update the dictionary/directory and the application programs. For example, if a field definition or coding scheme was changed in an active DED/D system the change would only have to be made once. The system would

THE DATA ELEMENT DICTIONARY/DIRECTORY AS A DATA MANAGEMENT TOOL

CHARACTERISTICS	APPLICATION	POTENTIAL BENEFITS
. Contains a unique identification, a set of physical characteristics, and a textual description for each data element	. Ordered listings of all entries or various classes of data with full or partial detail	. Simple and effective control of the data elements
. Shows relationships of elements to each other and to components of the system through programs and reports	. Cross-Referencing ability	. Reduction of data redundancy and inconsistency
. Specifies source, location, usage and destination of elements	. Finding a name from a description	. Enforcement of standards
. Validates and checks for redundancies	. Consistency and completeness checking	
. Contains security safeguards to control the accessibility to the data elements	. Generation of machine-readable data definitions	. Enforcement of security safeguards
		. Centralization of data elements as an aid in design and development of new systems
. Has a command language	. Extraction of Data Dictionary	
. Has reporting and retrieval capabilities	. Composite reports	. Consistency in documentation for data elements

EXHIBIT II-2

automatically modify the application programs. In a passive system the change would have to be coded into the DED/D and each application program affected by the change.

(c) Feasibility Studies

The objectives of coordinated feasibility studies in a total data management program are to facilitate review of intended program development so that redundant efforts and data inconsistency can be identified and efforts to improve coordination and reduce costs can be employed.

Feasibility studies are therefore required as a basis for determining and justifying the application of data management concepts within an organization's operational units and systems. During both the initial implementation and operational stages of a data management program, feasibility studies should be utilized to determine and justify appropriate levels of data standardization, required data element dictionary/directory (DED/D) contents and level of detail, and program policy and procedural requirements and impacts. Through the ongoing operations of the data management program, feasibility studies should be utilized to determine and justify new system development or system change efforts. Feasibility studies will not only correct data errors but assist in finding and correcting the reasons for the errors. These studies are an essential component of an ongoing data management program. Once study methodology is defined, future efforts will be simpler to adminisister.

(d) Quality Assurance Program

A common problem is a general lack of confidence in the quality of the data contained in an organization's systems. Sometimes this lack of confidence is not well founded, but is simply a matter of perception. In this circumstance an effective remedy is to educate the users on quality control measures being applied, and perhaps to publish statistics from time to time based on data quality audits.

However, in other cases, the lack of confidence is well founded and data contained in the systems is incomplete, inaccurate, or out of date. Reasons for poor data can usually be traced to the data entry point and such causes as:

- Lack of incentive on the part of those entering the data
- Confusion as to what data is required - content or form

- Poor data collection/entry forms design or poor procedures

- Lack of effective software edit capabilities.

In order to restore confidence in the data, and improve its quality, it is necessary not only to clean up data already in system files, but also to address the causes, such as those listed above.

Because of the variety of causes, any one or more of which may apply to a particular system, an audit is required to detect the problems. The objective of the audit is to identify the completeness and accuracy of the data in the files, and then to analyze procedures for data capture and processing to isolate the causes. The current files must be updated with good data and the capture and processing procedures modified to assure the future quality of the data. The steps that can be taken to improve the quality of the data include:

- Inclusion of effective edit/update software in systems -- proper edits will identify obvious errors such as data out of allowable ranges, letters in numeric fields, etc.

- Standardization of Data Definitions -- standard defintions eliminate confusion on data to be collected.

- Revised Forms and Procedures -- forms and procedures need to be easy to fill out and follow. Any ambiguity or unnecessary complexity should be eliminated.

- New Incentives -- The data collector must be given an incentive to collect data. The best method is to offer reports or data access techniques that are of use to the collector. If this cannot be done, then it is necessary to explore alternative means for collecting the data.

- Contractor Support -- It may be impossible to provide incentives to the current collectors for all or part of the collection and data entry process. If this occurs an alternative is to contract the work to private organizations. This is often done for data entry where incentive fees or penalties can provide the motive for accurate data entry.

The Quality Assurance Program must be an ongoing function with periodic audits conducted and improvements made. The Quality Assurance function is an essential part of effective information management, for without reliable data and the

confidence of its users, the most efficient and powerful system is worthless. Highlights of a data quality assurance program are shown in Exhibit II-3.

The quality assurance audit is also an essential tool for monitoring the utilization of other data management tools. Periodic review of existing systems and new system initiatives will aid in identifying:

- Inaccurate data or data deficiencies
- Poor procedures and lack of edit functions
- Non-compliance to data management policies and procedures.

Existing data standards, DED/D output, and program policies and procedures can serve as the basis for development of quality assurance criteria.

(3) Policies and Procedures

Policies are the rules set forth by the organizational unit responsible for the data management program. They include regulations on program responsibility assignments, approval flows, tools to be utilized, and quality assurance criteria.

Procedures describe the processes through which data management policies can be executed in a consistent manner. These include documented mechanisms for:

- Setting and promulgating data standards,
- Evaluating and approving requests,
- Auditing systems, and
- Updating data management tools.

Policies and procedures serve as a major data management program component since they provide the operational link between the organizational component and the program participants; and program participants, and the data management tools.

2. EPA CURRENT STATUS

This section presents the current status of data management in EPA. The current status is discussed in terms of existing policies and procedures for data standards, and the findings from review of a cross section of representative systems.

Data Quality Assurance

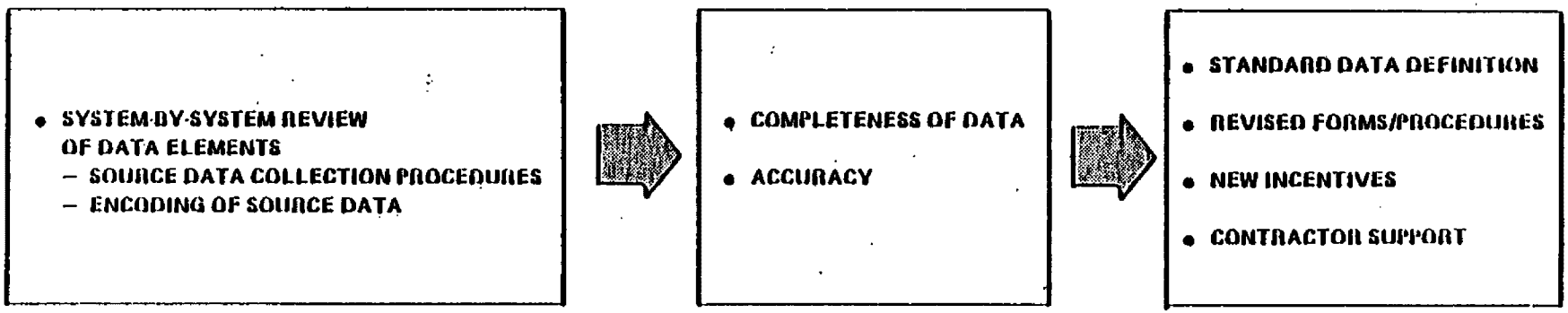


EXHIBIT II-3

(1) Policies and Procedures

There currently exists in EPA an ADP Manual which documents policies and procedures related to the utilization of data standards and approval processes for feasibility studies and system design and development. The Management Information and Data Systems Division (MIDSD) is recognized in the ADP Manual as the central area for coordination of data management efforts. The Manual states that,"All organizational elements in EPA, their contractors, and or grantees will promote the full utilization of Federal and Agency standard data elements and representations in the design and development of information systems". In addition any data elements and codes that are already in use by the Agency are to be adopted by the Agency as standards wherever it is practicable. Candidates for standardization can be recommended by any component organization of the Agency. Approved FIPS or Agency data standards are to be published and promulgated in the EPA Data Standards Catalog. Data element name, definition, item, coding scheme and code and abbreviation constitute prescribed catalog contents. Although an EPA-wide data dictionary initiative is currently underway, an EPA Data Standards Catalog does not yet exist.

In the current EPA Environment, the Management Information and Data Systems Division is to forward proposed standards to the Data Standards Coordinators in other organizational units for clearance. The Data Standards Coordinators are to coordinate standardization proposals within their areas and submit their comments on the standard to MIDSD. MIDSD then can resolve any conflicts in the proposed standard prior to publication of the approved data standard for EPA use. Initiation of potential data standards for review and approval is currently performed on a limited basis.

It is apparent from interviews with cognizant individuals within EPA and a review of the current level of data management and standardization activities that, although data management policies existed and are documented, they are not being actively implemented. There are many factors contributing to this circumstance including:

- Decentralized management of data and systems
- Little high level management awareness of the need for data management
- Limited resources in MIDSD

The policies and procedures are not currently accompanied by a dynamic program structure nor adequate tools for effective implementation and operation. This is further demonstrated by the system findings described in the next section.

(2) System Findings

Data standardization is a major data management tool which aids in improving the quality of data, increases the capability for data re-utilization, and simplifies system design and change efforts. Data re-utilization is the use of data collected for one system by a second system for perhaps a different purpose. The effect can be to reduce the redundant collection of data thus reducing the reporting burden on the public and private sectors. If system interfaces are developed, that can also lead to the reduction of storage of data in multiple locations.

EPA's need for data standardization was established on the basis of review of a cross section of fifteen active EPA systems which were selected by the EPA Advisory Committee. This review resulted in the determination of the types of data elements contained within sample systems. This analysis also identified the potential for information exchange or data re-utilization based on commonalities in naming conventions and coding schemes for data elements which were representative of the five data categories reviewed.

Exhibit II-4 presents a detailed systems category comparison matrix which lists data element name by category and by sample system. This exhibit shows the commonality of data elements in sample systems as well as the differences in naming conventions which impact data interpretations and transfer capabilities.

Exhibit II-5 presents a summary of findings for each data category. These findings are presented in terms of occurrences in systems reviewed, number of different elements, and number of different coding schemes for each data category. It is apparent from this more detailed perspective, that while data category commonalities provide candidates for data sharing, the proliferation of naming conventions and coding schemes prohibits or at least limits this potential.

The review of sample systems reinforced the concerns expressed by interview contacts in terms of limitations in the current environment relative to the application of data standards, limited application of system design and documentation standards, capabilities for data transfer, and current level of data quality. As a result of sample systems review, EPA's present status with regard to these issues can be summarized as follows:

(a) Data Standards

Elements with the same name having different definitions and different coding schemes are common throughout EPA Systems. Naming conventions for similar data elements also vary considerably. For example, in the Grants Information and Control System (GICS) the city name refers to the name

Sample Systems Comparison Matrix

	COS	EDS	ERSS	FAS	GICS	NSIS	AEROS NEDS	NEEDS	PEAS	PCS	PPIS	RAAS	AEROS SAROAD	STORET	TOXICS
I. FACILITY IDENTIFIERS															
COMPANY/AUTHORITY NAME		x	x			x		x			x				x
COMPANY/AUTHORITY CODE		x	x	x				x			x				x
FACILITY NAME	x	x	x			x	x	x	x	x		x			x
FACILITY ID NUMBER	x	x	x		x	x	x	x	x	x	x	x		x	x
II. MONITORING SAMPLE STATION SITE															
SITE NAME						x							x		
SITE CODE	x	x					x						x	x	
III. GEOGRAPHICAL LOCATION															
REGION CODE	x	x	x		x	x			x	x	x		x		x
STATE CODE	x	x	x		x	x	x	x	x	x	x	x	x	x	x
STATE NAME		x									x				
COUNTY CODE	x	x				x	x	x		x		x	x	x	x
COUNTY NAME	x	x			x		x	x		x			x		x
CITY CODE	x	x					x			x		x	x	x	x
CITY NAME	x	x	x		x	x	x	x	x	x	x		x		x
ADDRESS	x	x	x		x	x			x		x	x	x		x
ZIP CODE	x	x	x		x	x		x			x				x
OTHER GEO.	x	x		x	x	x	x	x	x	x		x	x	x	
IV. PARAMETER UNIT IDENTIFIERS															
UNIT		x	x			x	x			x			x	x	
V. QUALITY ASSURANCE CODES															
Q.A.	x					x	x		x						x

SUMMARY OF FINDINGS

DATA CATEGORY	OCCURRENCE IN SYSTEMS REVIEWED	NUMBER OF DIFFERENT ELEMENTS	NUMBER OF DIFFERENT CODING SCHEMES	NOTES
FACILITY IDENTIFIERS	83%	10	16	VARIETY OF DEFINITIONS
SAMPLE MONITORING SITE	47%	0	3	
GEOGRAPHICAL LOCATION	100%	36	20	FIPS CODES USED IN FOUR SYSTEMS
PARAMETERS UNIT IDENTIFIERS	47%	2	6	THREE SYSTEMS HAVE SAME CODING SCHEME
QUALITY ASSURANCE CODES	33%	6	3	NO METHOD SET UP FOR CODES

EXHIBIT II-5

of the city in which the headquarters of the applicant is located, however, in the NEEDS system the city name is the name of the city or town in which the facility is physically located.

The National Bureau of Standards (NBS) has been committed to a policy of adopting voluntary ADP standards developed by the American National Standards Institute for use in the federal government. There is not uniform adherence to these standards in EPA systems.

EPA is currently engaged in several projects that deal with the development and implementation of data standards. The Inter-agency Regulatory Liaison Group (IRLG) Common Codes Project is a good example. The objective of this project is to identify specific data elements within selected key data element categories, and recommend preferred naming conventions and coding schemes for the elements. The proposed system of common codes would be applied to new as well as existing information systems which are of common use to the four member IRLG agencies.

Data standardization is also being undertaken at a programmatic level. The Wastewater Treatment (WWT) Program for the Office of Water Program Operations (OWPO) is currently engaged in this type of initiative. The WWT data management and standardization program will utilize tools such as cross reference files, standard definitions and a data element dictionary/directory.

While data standardization attempts can be facilitated by aggregating and analyzing the attributes of data components in systems, there is currently no central location for metadata within EPA to afford this capability. Some offices maintain their own data inventories, but there is no central data reference source such as a data element dictionary/directory. There is, however, a current study in which a common facility identification coding scheme for facilities monitored by the EPA is being developed. The resulting facility file can serve as input to an EPA-wide data dictionary/directory system which is currently under consideration.

(b) System Design and Documentation Standards

On an EPA-wide basis, limited consideration has historically been given to the need for data management or standardization with respect to systems design and documentation for proposed systems. Some offices have established and enforce system design and documentation standards on their own, however, the Agency as a whole does not strictly enforce established standards. The impending potential for massive

conversion following the 1980's ADPE procurement has emphasized the need for documentation standards. There is an effort by MIDSD currently underway to develop new Agency-wide standards for system documentation. Upon preliminary review, this appears to be a particularly practical approach in that it relates the rigor of the standard required to the size, life expectancy and scope of the system.

(c) Data Transfer

There are two forces in EPA that are leading to increased interest in the sharing of data between offices and among Headquarters, Regions and States. The first is the increased interaction between programs (e.g. construction grants and NPDES permits; water enforcement and WWT facility operation and maintenance; and multi-media new source tracking.) The second force is the increased delegation of programs to States by Regions. Current lack of standardization presents difficulties in sharing data across systems, as well as data interpretation and data quality. Improved mechanisms are required to aid EPA Headquarters and Regions in the effective management and utilization of data.

(d) Data Quality

Interview contacts consistently expressed a lack of confidence in the quality of available data. Audits of individual system files provide documented evidence of incomplete data, and data inaccuracies. Not only does poor quality impose a manual burden on generation of major reports in assuring accuracy and completeness, but there is the potential for reporting inaccurate data to other Federal regulatory agencies or to the public. There is not currently an active Agency-wide quality assurance program so that any activity is left up to the initiative of individual system managers or program offices. When activities are undertaken, they are usually crisis oriented directed at merely cleaning up existing data rather than correcting the causes for poor quality. There is currently a Blue Ribbon Monitoring Group in EPA, chaired by Richard Dowd, which is studying the many aspects of monitoring data. One of the major considerations of this group is the quality of this monitoring data and the confidence level of the data that is available.

3. OVERALL EPA REQUIREMENTS

A data management and standardization program for EPA is directed at several objectives which will facilitate management of environmental programs and help to maintain a positive relationship between the Agency and the public. Key among these objectives are:

- Improved data quality to provide a basis for sound decisions, to minimize the use of scarce personnel resources for data collection and validation, and to permit efficient use of automated capabilities

- Facilitated data transfer to permit effective communication between related programs, accurate aggregation of multi-media data, and sharing of capabilities to reduce redundant system development

- Reduced reporting burden on industry, the states, and the Agency's own regions so that a positive relationship can be maintained and resources of all entities can be applied more efficiently.

- Development of a focal point for answering inquiries to respond to the increased visibility of the more mature programs in the Agency. This will help to assure accurate and complete response to requests for information.

Current capabilities in the Agency do not adequately support these objectives. It is concluded therefore, that a need does exist for a data management and standardization program for EPA. This conclusion is based on the fact that both perceived and documented benefits of data management and standardization provide potential for resolving current data accessibility and information exchange problems experienced by EPA. Since there is a need to share data across systems, it is important that these systems adhere to the same standard data definitions, coding schemes, and documentation techniques. The systems that currently support the programs in EPA are under the control of the various offices which support programmatic functions. Enforcement of standards across these systems therefore requires coordination between Programmatic areas. A data management program can provide the mechanism to permit the required data sharing. This program can also provide a medium for the coordination which is required to plan for new systems or enhancements to existing ones.

Because data management involves all aspects of data, from its collection to its final transformation into information products, it offers the prospect of enhancing those products in several ways. First, improvements in the validity, reliability and timeliness of data will be directly reflected in the corresponding information products. Second, information products viewed apart from their respective systems may become candidates for standardization, and standardization may make the information both more meaningful and more desirable. Finally, when information products are viewed across the Agency, there arises the possibility of two or more such products being recognized as complementary. Actions may then be taken to enrich the products in some manner of combination.

As described in Section 1, the component of a data management program are policies and procedures, tools, and a data management

organizational structure. Alternative approaches for policy and organization are described in Chapter III. Highlights of the tools required for EPA are described in the paragraphs which follow.

(1) Standards

The development, maintenance and rigorous application of standards are required for:

- data elements
- data acquisition
- system design and documentation.

There is a clear need to standardize data elements to permit the necessary sharing of data between related programs and to help improve data quality. The standardization may be limited to key data elements which serve as linkages to related systems, or could be targeted for the entire Agency base of data. A high priority candidate for standardization, for example, is the facility or authority identification data category which is a common element in the majority of the Agency's larger systems. A primary mechanism for effecting data element standardization is the data element dictionary/directory described in Section 1.

There is also a current need in EPA to enhance data acquisition procedures. Enhancements to the current approval process for reporting requirements could include:

- Verification of standard names, definitions, and coding schemes,
- Combination and reduction of the number of data collection forms,
- Re-utilization of data already collected
- Verification of the inclusion of quality assurance elements,
- Verification of proper economic analysis to measure the cost-effectiveness of data collection techniques.

Data acquisition processes should involve the Regions as well as Headquarters. Since the Agency requires an increasing amount of data from the States and Regions, it is necessary to establish and enforce standard methodologies and procedures for data acquisition. This will help to assure quality and compatability of the data received at headquarters.

Standard procedures should be developed and followed for all stages in systems development. The approval process for system development must include an in-depth look at the system software in order to evaluate the possibilities of using standardized packages or subroutines. Any improvements or modifications to the existing hardware or software should be approved through the data management program. Without such approval, systems would be modified using non-standardized methodologies resulting in little or no possibility for data transfer or interface.

Policies and Procedures for the standardization of system documentation also need enhancement and enforcement. These policies should specify the types of attribute data to be included as well as basic formats for the documentation, which should be user-oriented and easy to understand.

Standard programs and subroutines, such as conversion from English to metric units of measure, should be catalogued during the approval process. This would provide an inventory of programs, and would reduce any programming duplication. These programs could be catalogued in a Data Element Dictionary/Directory (DED/D)which would serve as a centralized systems reference source. The DED/D would use its capability as a copy library in cataloging these standardized programs and subroutines.

(2) Data Element Dictionary/Directory

Due to the increased visibility of EPA, an increasing number of inquiries are coming in from various entities such as:

- Congress
- States
- Public (FOI)

Thus, an increasing amount of resources is needed to respond to this ever-increasing amount of inquiries.

Currently there is no mechanism available in EPA that facilitates timely reference of available data resources. EPA therefore needs a central repository of information that can be used to locate data relevant to any program within the Agency. This requirement can be satisfied by establishing a centralized Data Element Dictionary/Directory. The Dictionary should contain a list of all key elements relevant to the EPA programs, along with a definition of each element including any special coding schemes. The Data Dictionary can be supplemented by a Data Directory which will meet identified requirements to establish a mechanism which provides information on how data can be accessed. The Data Directory will identify location of data by

systems. This combination of a data dictionary and directory will not only provide a focal point for locating data, but provides a consolidated reference source which can be accessed to determine standards utilization. This will aid in the identification of standardization candidates, or enforcement of existing standards use.

(3) Quality Assurance

Quality assurance is another important tool in a data management program. The prevalent lack of confidence in data in Agency systems clearly demonstrates the need for such a program. A quality assurance program must be directed at not only cleaning up existing data but also at removing the causes of poor quality. These causes can be identified through system and procedural audits, and some can be removed through improved data entry forms, comprehensive data input software edits, and clarification of data element definitions. However, a major factor in poor quality in the Agency is the circumstances which require the Regions to report data to Headquarters which is not data which the Regions require for their own operations. When this circumstance truly exists, imaginative alternatives to acquiring data must be considered. For example:

- The Regions may be helped to discover the usefulness of the data to themselves as well as to Headquarters
- Contractors may be used to enter the data under direct responsibility of Headquarters
- Sample data collection may be determined to be sufficient rather than Agency-wide data collection on a recurring basis.

Since an effective quality assurance program does require significant funds, an efficient plan for addressing the program which emphasizes long-lasting solutions and self-correcting procedures must be developed and management commitment obtained.

These requirements form a basis for the data management and standardization program for EPA. The next chapter describes alternative strategies for statisfying these requirements.

Alternative Analysis

III. ALTERNATIVES ANALYSIS

III. ALTERNATIVES ANALYSIS

In this Chapter, Arthur Young & Company presents the analysis of alternatives for the major components of a Data Management and Standardization Program in EPA. An effective EPA-wide data management and standardization program must consist of three major components:

- an organization structure which supports program implementation and administration;
- data management tools such as standards, data element dictionary/directory (DED/D), feasibility studies, and quality assurance programs which produce program products; and
- policies and procedures to govern program operations.

Each of these components is discussed in the following sections. Where alternative approaches exist, the discussion will focus on the alternatives' attributes and potential effects.

1. ORGANIZATIONAL STRUCTURE

EPA requires a somewhat different approach to the development of a Data Management and Standardization Program than is advocated in the text books. The advocated approach is to define the requirements for the tools, develop the policies and procedures that will best implement and control the tools and then define the organization to administer the policies and procedures and operate the tools. The formal organization in EPA requires a modification to this approach because certain alternatives for a data management program framework are not possible. Specifically, the multi-faceted nature of the Agency's programs as well as the line-staff relationship between Headquarters, Regions, and Laboratories mandates division of responsibility that allows only coordination, not dictation of efforts. The effects of the Agency structure on each alternative will be discussed along with the merits of each alternative.

Traditionally ADP management has looked at the alternatives for control as being either centralized, decentralized or distributed. These concepts have been primarily applied to computer management. The concept in favor has been dependent on the state-of-the-art in hardware. Large computers with advances in telecommunications and remote access equipment led the move from decentralized to centralized computing facilities and control. Recent advent of mini and micro processors with further advances in telecommunications is leading the move towards distributive systems.

The principle behind each control concept are:

- Centralized - authority and responsibilty, being placed at the highest common point in the organization to effect maximum control

- Decentralization - authority and responsibility being located as close to the level of the organization where the actual work takes place to speed the decision-making process.

- Distributive - locate the authority and responsibility at the level in the organization that is best suited to perform the specific task.

The same trends are being followed in the recently recognized area of information resource management. The concepts are the same: decentralized, centralized, and a form of distributive that we have named hierarchical. Our discussion of each concept is placed in terms of the EPA structure. Exhibit III-1 presents a graphic overview of the three structures as they might apply to EPA.

Included in the discussion of the alternatives is the evaluation of the alternatives potential and impact in EPA. In preforming the evaluations we have considered the inherent ability of each alternative to obtain:

- Top management commitment

- Programmatic participation

- Balance of authority and responsibility

- Control and coordination of EPA information resources.

It should be noted that, in the discussions of the concepts of Decentralized and Centralized Organizations, we are looking at the extremes of the spectrum. Actual implementation is typically less stringent.

(1) Decentralized Organization

In the decentralized organization structure, data management responsibility would be assigned at the programmatic level under the control of the Deputy Assistant Administrator who has primary responsibility for a particular program. Program management would be controlled by an Oversight Committee that would coordinate with ADP support staffs for the program components to assure that data management program concepts were properly applied throughout programmatic level operations. The Oversight Committee would be responsible for the setting and promulgating of program policies; review of feasibility studies, system designs and implementations; oversight audits to assure standards compliance;

ALTERNATIVE ORGANIZATIONS

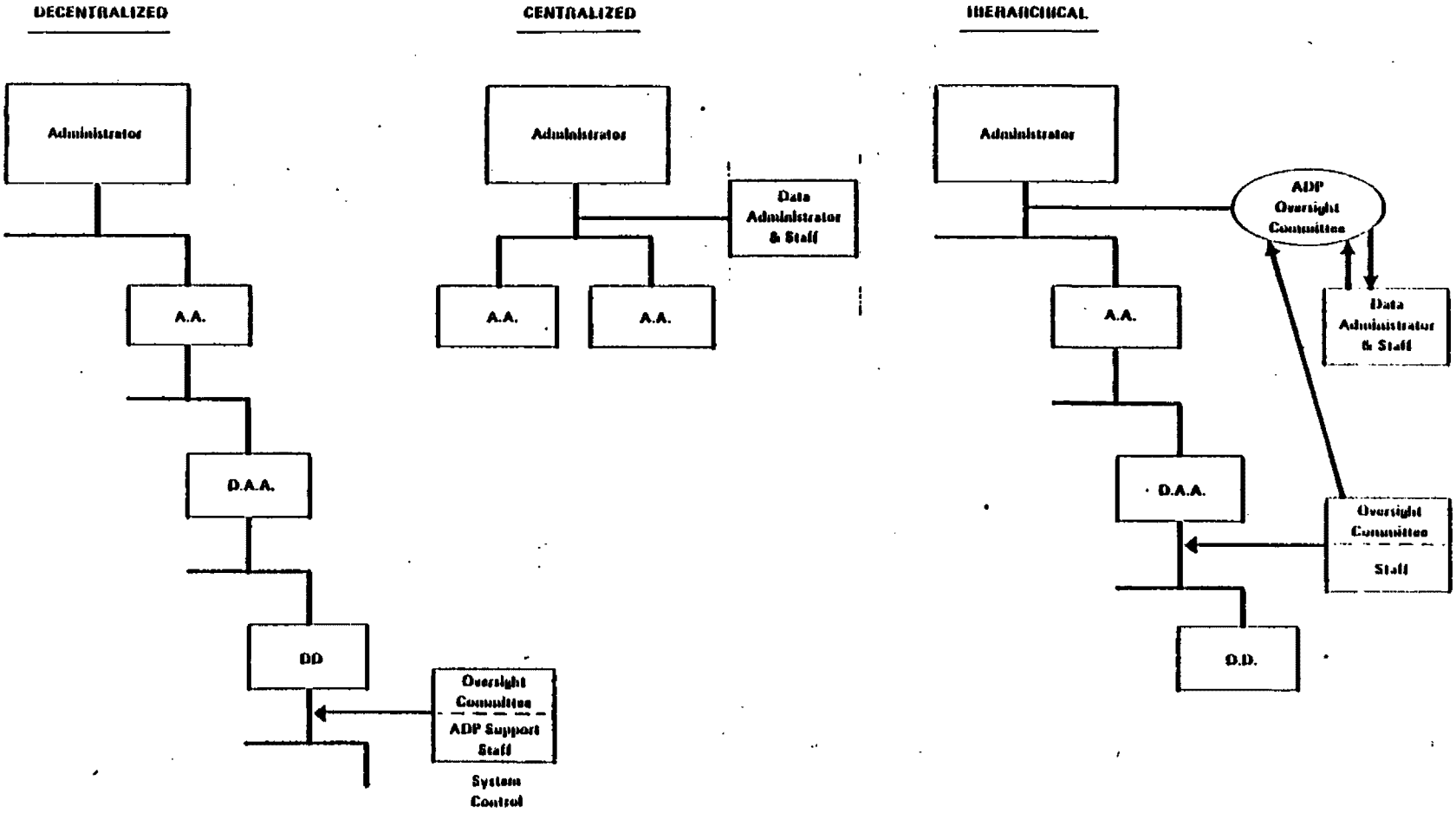


EXHIBIT III-1

and system coordination to assure that programmatic requirements are met.

The ADP support staff, which may be one centralized group for the program or separate staffs for each system supporting the program, would be responsible for the performance of feasibility studies and systems design and implementation. In addition, they would be responsible for updating data element detail in individual systems catalogues and/or DED/D. Systems groups would also be responsible for performing systems audits to assure data quality, and system evaluation to assure that system requirements were being met.

As can be seen in Exhibit III-2, an overview of the intersystem life cycle data management for this alternative organization structure, no EPA-wide management of the data resource is included in this alternative. With the inclusion of some MIDSD oversight of the systems, this alternative approximates the current status of information processing in most of EPA.

This alternative does not support EPA-wide requirements for data management and standardization since program activity and control is confined to the programmatic level. While improvements to data quality and accessibility may be obtained on a programmatic basis, the benefits to be accrued from data management and standardization are limited to the degree of program implementation which is chosen to meet the requirements of individual programmatic units. As an organization, EPA's need to share data, minimize system development and change efforts, improve overall data quality, reduce reporting burdens, and meet external reporting or information exchange requirements cannot be met. This alternative is therefore infeasible for use because it does not meet the defined criteria for an effective EPA data management and standardization program.

(2) Centralized Organization

In the centralized organization structure, all EPA-wide data management responsibility would be assigned to a Data Administrator (DA) and staff which reports directly to the Administrator. The Data Administrator would be responsible for the setting and promulgating of program policies and the review of feasibility studies, and system design and implementation. The Data Administrator's staff would be responsible for monitoring the use of program policies; performing feasibility studies and system design and implementation, or managing contractors who perform the projects; updating the EPA-wide data element dictionary/directory; auditing systems for standards compliance and data quality; and conducting system evaluations to assure that both EPA-wide and programmatic requirements are met.

DECENTRALIZED ALTERNATIVE
INTERSYSTEM LIFE CYCLE DATA MANAGEMENT

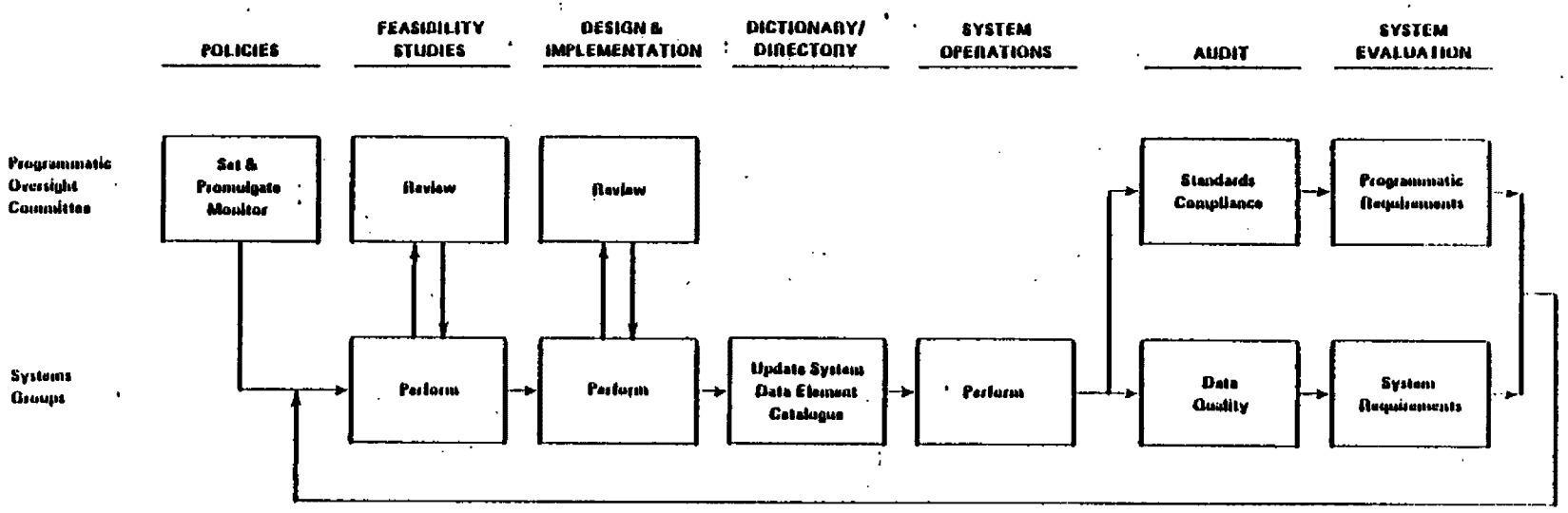


EXHIBIT III-2

Under this alternative organization structure, the responsibilities of the programmatic staff would be limited to performance of system operations. Exhibit III-3 on the following page presents an overview of the intersystem life cycle data management for this alternative organization structure.

This alternative supports almost all the EPA-wide requirements for a data management and standardization program. It does not, however, support the need to encourage active participation by programmatic areas. In fact, current programmatic level data management responsibility is reduced to maintenance of ongoing system operations. This alternative prescribes a complete reorganization of EPA-wide data control responsibilities and the creation of an additional Data Administrator position and supporting staff. The newly formed data administration function requires high-level, Administrator staff positioning to give legitimate power to the strong role of the Data Administrator required in this alternative.

The logistics for the implementation and ongoing operation of this alternative program strategy are both complex and contrary to EPA's current programmatic oriented organization structure in both the Regions and Headquarters. The effect is intensified by the relative independence of the Regions. Since programmatic and Regional support of a data management and standardization program is essential to program success, an alternative which does not facilitate and encourage such support may never achieve its potential benefits due to lack of participation and cooperation. This alternative is therefore infeasible for use since it does not have the potential for effective implementation in the current EPA organizational environment.

(3) Hierarchical Organization

The hierarchical organization structure provides for both an ADP Oversight Committee which reports to the Administrator, and individual oversight committees for each program. The ADP Oversight Committee which reports to the Administrator would be responsible for the setting and promulgating of program policies. A Data Administrator and staff would assist the ADP Oversight Committee in the documentation of policies and development of supporting procedures. The Data Administrator and staff would also support the ADP Oversight Committee by performing review functions for feasibility studies, and system design and implementation; updating key elements in the EPA-wide dictionary/directory based on input from programmatic areas; perform system audits to assure standards compliance; and conduct system evaluations to assure that EPA-wide requirements were being met.

The Programmatic Staff would retain current responsibilities for feasibility studies, system design and implementation, and

CENTRALIZED ALTERNATIVE
INTER SYSTEM LIFE CYCLE DATA MANAGEMENT

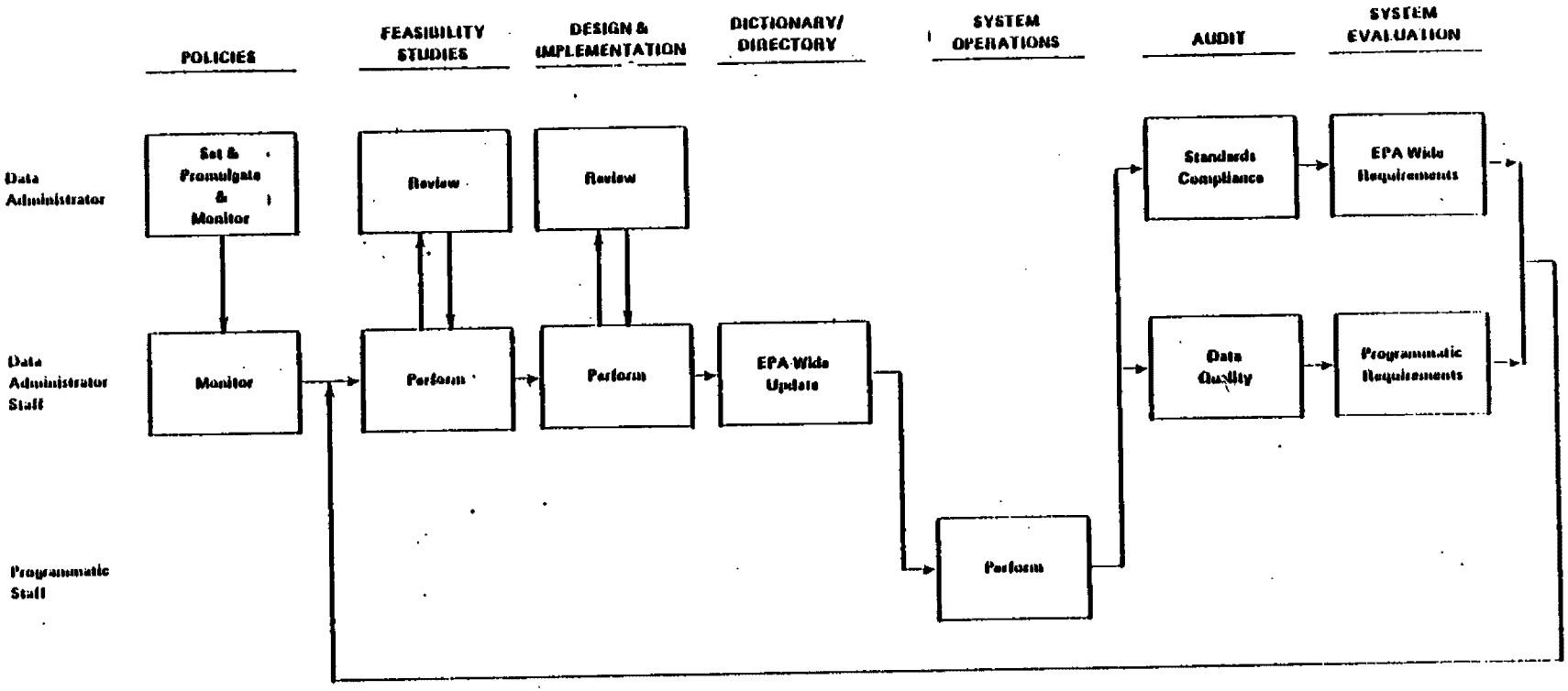


EXHIBIT III-3

system operations. In addition, they would be responsible for maintaining a detailed program specific DED/D and providing the DA's staff with the metadata needed for the Agency-wide dictionary/directory. (Details on the DED/D alternative follows in the Tools Alternative section.) Programmatic areas would also be responsibile for auditing data quality and performing system evaluation from the perspective of programmatic requirements.

An oversight committee would be formed for each programmatic area to monitor the adherence to programmatic level data management concepts and provide valuable input to the EPA-wide ADP oversight committee in terms of additional policy and procedures requirements and data management program operations. Exhibit III-4 on the following page presents an overview of the intersystem life cycle data management for this alternative organization structure.

This alternative supports the EPA-wide requirements for a data management and standardization program. Responsibilities for data management are assigned on a hierarchical basis so that program objectives are properly addressed by the levels of EPA management which can best assure and control that these objectives are met. Participation is encouraged from programmatic areas through oversight committees which are represented on the EPA-wide ADP Oversight Committee. The ADP Oversight Committee then reports on program activities to th Administrator. A Data Administrator and staff assist with required EPA-wide review processes, and monitor the administrative aspects of the program.

In this alternative strategy, both EPA-wide, programmatic, and regional level data management and standardization needs are accomplished with limited changes to current organization structure or responsibility assignments. This alternative will also provide one focal point for initial contact with external agencies. In this context, the program will provide an effective mechanism for interagency information exchange such as that anticipated by the IRLG common codes effort. Already the OWPO/WWT long-range plan is establishing the first programmatic/office sub-groups which provides a pilot implementation of data management concepts for the agency which can effectively feed into an EPA-wide program implementation.

This alternative program implementation strategy therefore appears to provide the most effective data management program implementation strategy for the current EPA organizational environment. In some areas EPA is already developing organizations which incorporate many of the data management and standardization concepts contained in the above alternative program implementation strategies. What is needed, however, is a coordinated program implementation effort. How this coordination can be achieved, to what degree it must be employed, and the speed at which it can be done can be further assessed

HIERARCHICAL ALTERNATIVE
INTERSYSTEM LIFE CYCLE DATA MANAGEMENT

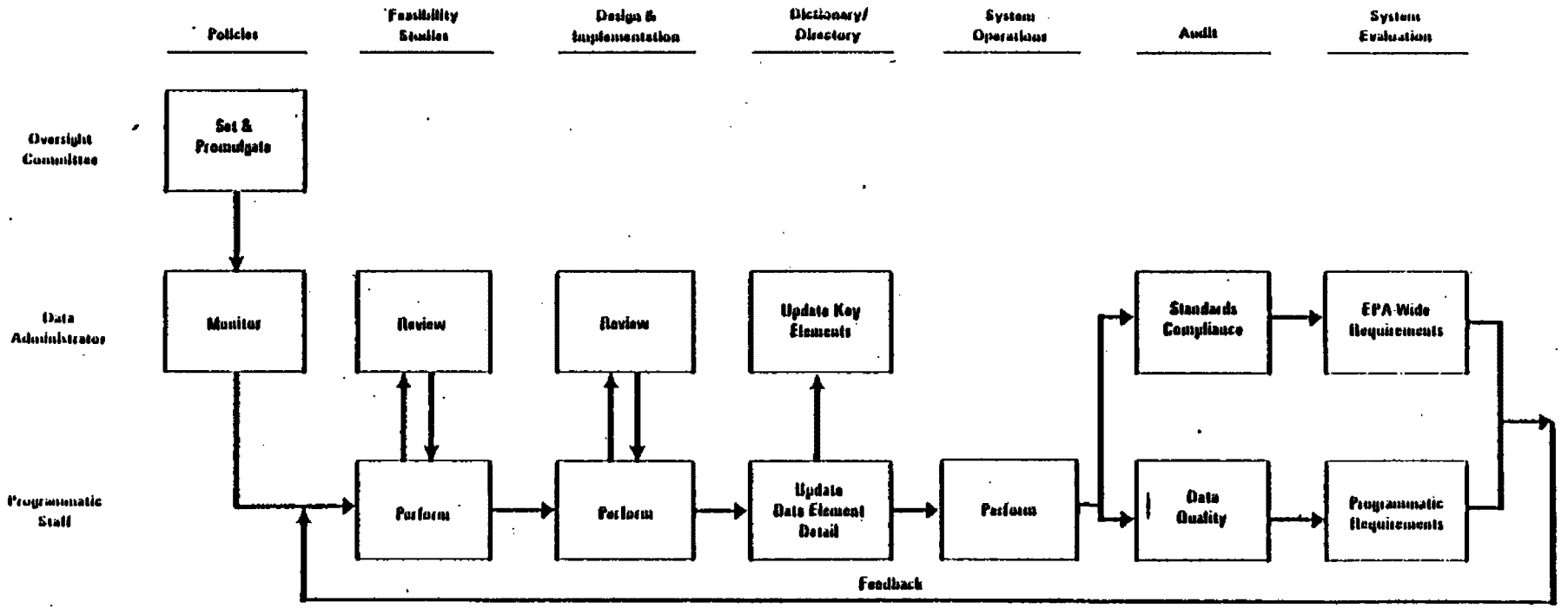


EXHIBIT III-4

based on findings of the Nolan, Norton and Company study to determine EPA's current position in the organization-wide system development life cycle.

2. DATA MANAGEMENT TOOLS

In Chapter III we identified the major tools and their characteristics that are required for an effective data management program in EPA. These tools already exist in some form in the Agency or are currently under development. The alternatives in regard to the tools are primarily in the delegation of responsibility and authority for the different phases of the tools' development and application. The degree of delegation is dependent on the organization structure selected. Exhibit III-5 presents the recommended degree of delegation for each structure.

One set of alternatives that should be addressed with regard to EPA data management tools is the approach for the Data Element Dictionary/Directory (DED/D). The pertinent alternatives deal with:

- Centralized vs Hierarchical DED/D
- Active vs. Passive DED/D

The considerations for each of the alternatives are discussed in the following paragraphs.

(1) Centralized vs. Hierarchical DED/D

There are a number of data dictionaries and dictionary/directories currently in use in EPA. Most of them are systems oriented like GICS and STORET. There are several program or multiple system dictionaries available such as AEROS in RTP. This needs to be a consideration in evaluating the alternatives. The first option is to retain these dictionaries with their various formats, contents, and media and to overlay an Agency-wide directory. The Agency directory would be organized by data category and contain only key identification data, or metadata, on data elements in the programatic data bases. The type of information that might be carried in the Agency directory would include:

- Data element name with brief definition
- Programmatic DED/Ds that have detailed metadata on element
- Systems on which the element is carried
- Contact name and phone number for both relevant systems and DED/Ds.

ALTERNATIVE COMPOSITION OF THE TOOLS AND POLICY AND PROCEDURE RESPONSIBILITY

TOOL \ ALTERNATIVE	DECENTRALIZED	CENTRALIZED	HIERARCHICAL
STANDARDS	• EPA-Wide Guidelines • System Groups Sets and Enforce	• EPA-Wide Standards – Set by Data Administrator (DA) – Enforced by DA Staff	• EPA-Wide Standards Set by ADP Oversight Committee • Enforced by DA Staff
DATA ELEMENT DICTIONARY/DIRECTORY	• Data Element Catalogue for Each System • Maintained by the System Groups	• EPA-Wide Data Dictionary/Directory • Maintained by DA Staff	• EPA-Wide Key Element Dictionary/Directory – Maintained by DA Staff • Programmatic Data Dictionary/Directory – Maintained by Program ADP Coordinator & Staff
FEASIBILITY STUDIES	• System Groups Conduct • System Perspective • Oversight Committee Review	• DA Staff Conducts • EPA-Wide Perspective	• Programmatic Staff Conducts • DA Staff Reviews for EPA-Wide Perspective
QUALITY ASSURANCE	• Systems Groups Conduct	• DA Staff Conducts	• Systems Groups Conduct Data Audits • DA Staff Review Systems for Compliance to Standards

EXHIBIT III-5

The objective would be to use the Agency directory as a central location to direct inquiries to the individuals and system that can best answer the questions. For example, if a congressman requested the number and dollar amount of waste treatment plants under construction in his district, the Agency directory would direct the inquiry to the proper person or persons in the OWPO, Grants Administration and/or the Regions who, according to the directory, would have the information. Each systems/program would be responsible for maintaining their own DED/D and supplying relevant metadata to the group charged with maintaining the Agency directory.

The second alternative is to develop a completely integrated Agency-wide DED/D. MIDSD has initiated a pilot project, with contractor support, to test this concept. The advantage of this approach is the central location of all metadata that facilitates the quick location of information, the monitoring of data definitio and coding scheme standards, and the uniformity DED/D content. The disadvantage is that a Agency-wide DED/D will be massive in size which will require a major effort to maintain. Thus, potential for problems is compounded when the maintenance is done by people not familiar with the actual systems. In addition, under this approach the programmatic areas lose proximity to the DED/D. Experience has shown that the program areas are more committed to capabilities which are directly under their control. The effect may be that changes in systems that require DED/D updating will not be reported.

(2) Active vs. Passive DED/D

The primary difference between an active and passive DED/D is that the active DED/D interfaces with the programs to simplify the maintenance of both the DED/D and the application programs. A change in either field description or coding schemes is entered into the DED/D then the various application programs that use the file and the coding schemes are automatically updated. This is a sophisticated firmware package concept that will require significant modifications to all current programs. The advantages of the active DED/D in reduction of maintenance and assurance of standards in system development are such that it should be seriously evaluated for inclusion in the 1980's Procurement. We do not recommend the active DED/D at this time because of the massive effort that would be required to modify the application programs to permit the use of an active DED/D. In addition, for an active system to be fully useful, standard data definitions and coding schemes need to be in effect. The 1980 ADPE Procurement affords an excellent opportunity for implementation of both comprehensive data standards and an active DED/D because a major conversion effort is anticipated to operationalize the current programs on the new hardware. While the programs are being modified for the new operating system, the incremental cost for modifying or including source code to effect standards and the

active DED/D will probably be more than offset by the benefits to be gained by each. A study should be performed to test this assumption.

Our recommendation for the DED/D is to implement a passive dictionary/directory on a hierarchical basis. This allows the utilization of the current DED/Ds in operation in EPA with the development of a centralized directory to be used to point to specific data elements in the detailed programmatic/system DED/D. Standards should be set for the development of additional DED/D. The current MIDSD effort to build a DED/D for selected systems is the vehicle for not only setting standards but for eventually providing DED/D capabilities to the other programmatic areas. Finally, we recommend a detailed feasibility study for the implementation of an active DED/D in conjunction with the 1980 ADPE Procurement. This study should be conducted relatively soon so that the ADP specification for supporting an active DED/D can be included in the RFD for the Procurement, if the study deems the system feasible.

3. POLICIES AND PROCEDURES

As stated earlier, the policies and procedures needed to implement and operate a data management program are dependent on the level of information management desired, the delegation of responsibility and authority throughout the agency, and the exact nature of the tools to be implemented. The development and implementation of policies and procedures must be of primary priority to the ADP Oversight Committee.

4. RECOMMENDATION

The specific recommendations in regards to the tools used in the Data Management Program are not new to EPA. Most, if not all, are in one stage or another of development or operation currently in sections of EPA. They range from pilot efforts that have been initiated by EPA since the inception of this project to fully active programs that have been successfully in existence for years. What this project has accomplished is to identify these isolated programs and to prepare a framework for coordinating and promulgating these efforts throughout EPA.

(1) Organizational Structure

The organizational structure for the data management program, with its corresponding responsibilities and authorities, is the basis for pulling together an effective program in EPA. In performing the evaluation of the alternative organizational structures, we had to consider the inherent ability of each alternative to obtain:

- Top management commitment,

- Programmatic participation,
- Balance of authority and responsibility, and
- Control and coordination of Agency-wide information resources.

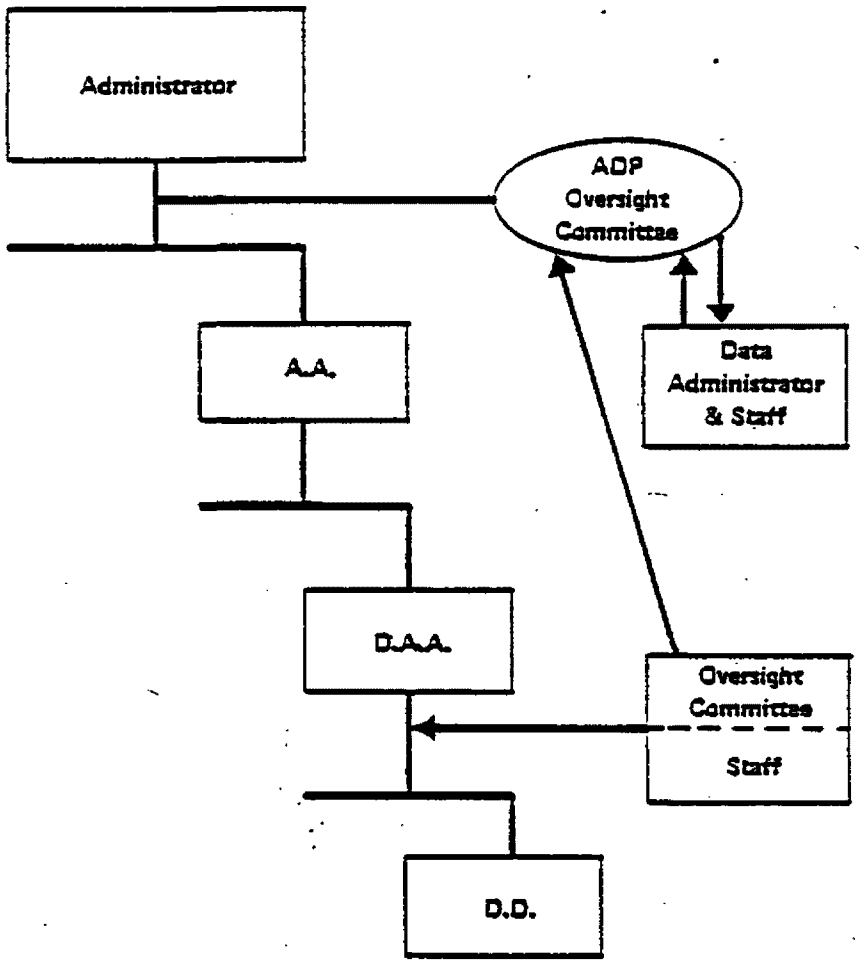
Our recommendation is a hybrid of the classical centralized and decentralized management approaches that we call hierarchical. The basic philosophy is analogeous to distributive processing in computers - locate the work at the level in the organization best suited to perform the specific tasks.

The hierarchical organization structure provides for both an ADP Oversight Committee which reports to the Administrator, and individual oversight committees for each program. This concept is presented graphically in Exhibit III-6. The ADP Oversight Commitee which reports to the Administrator would be responsible for the setting and promulgating of program policies. A Data Administrator (DA) and staff would assist the ADP Oversight Committee in the documentation of policies and development of supporting procedures. The Data Admnistrator and staff would also support the ADP Oversight Committee by performing review functions for feasibility studies, and system design and implementation; updating key elements in the EPA-wide data element dictionary/directory based on input from programmatic areas; perform system audits to assure standards compliance; and conduct systems evaluations to assure that EPA-wide requirements were being met.

The Programmatic Staff would retain current responsibilities for feasibility studies, system design and implementation, and system operations. In addition, they would be responsible for maintaining a detailed program specific data element dictionary/directory and providing the DA's staff with the data needed for the Agency-wide dictionary/directory. (The data element dictionary/directory concept is described in Section 3. Details on the DED/D alternative which follows.) Programmatic areas would also be responsible for auditing data quality and performing system evaluation from the perspective of programmatic requirements.

An oversight committee would be formed for each programmatic area to monitor the adherence to programmatic level data management concepts and provide valuable input to the EPA-wide ADP oversight committee in terms of additional policy and procedures requirements and data management program operations. Such a programmatic oversight committee is being formed by OWPO to support the Wastewater Treatment (WWT) Facility program. We recommend that the chairperson of the programmatic committee be a member of the Agency Committee.

Hierarchical Organization Structure



(2) Data Management Tools

We recommend that priority and support be given to the current projects that are developing individual tools for specific programs in EPA. These projects should be viewed as pilots upon which an Agency-wide program can be built. With this perspective it is necessary that the projects be monitored to assure that they meet the guidelines and objections of the tools as described in Chapter II. The specific action that should be taken for each tool is discussed in the following paragraphs in the context of the recommended organization.

(a) Standards

There are efforts currently underway in each area of standards that must be supported. They need to be coordinated such that they are ready for Agency-wide implementation before or during the 1980's ADPE Procurement conversion.

Data Standards - There are a number of efforts going on in EPA dealing with data definitions, coding schemes, and naming conventions. They include:

- MIDSD Data Element Dictionary/Directory Project
- IRLG Common Codes Project
- Interagency Toxic Substances Data Committee
- Region 2 Facility File Pilot Project
- OWPO Long-Range ADP/IS Plans

They need to be coordinated to assure that the codes and definitions established in one effort are compatible with the other projects results. The IRLG, Region 2 and OWPO projects are all dealing with facility/authority/establishment identifiers which must result in the same or compatible coding schemes.

Our sample of systems and projects in EPA was not exhaustive. It is very likely that there are other efforts going on to develop special coding schemes and definitions for specific program areas. A study needs to be instituted to identify all these projects and to identify those projects developing codes and definitions for the same data categories. Efforts should then be made to coordinate the development. In addition, other data categories not currently being addressed, but that should be standardized need to be identified and work begun on their standarization so that they are ready for the 1980 Procurement Conversion. As noted in Chapter II, not all data elements used to be standardized.

What may be needed is a method of relating data, such as cross-reference indices or conversion tables for measurement units.

System Design and Documentation Standards - An effort to prepare new standards is being conducted by MIDSD in preparation for the 1980 Procurment. We did not review this project in detail, but the concept of having three (3) levels of detail documentation dependent on the use and complexity of the system is an excellent idea that has worked extremely well in other agencies. This effort should be enthusiastically supported by top management with each AA assuring that his area meets the standards by the 1980 Procurement Conversion. Not only is this good on an Agency level but it will help minimize the description and potential errors from ADPE conversion that could seriously impact the operations of the programs under each AA. It is in their own best interest to assure that the standards are met by conversion time.

Data Acquisition Standards - This is the area where we found the least activity. There are isolated efforts addressing some aspects of data acquisition. For example:

- Forms redesign - Operations and Maintenance in OWPO
- Review of reporting impact of new requests on Regions - Program Reports Division
- Coordination of OMB approval for new forms - Program Reports Division
- Evaluation of document handling and tracking - OTS and OPP
- Data Transfer between COE and Regions, Regions and Headquarters, WWT program and permits - OWPO Long-Range ADP/IS Plan.

These efforts are addressing individual areas that are a problem to the specific groups. A coordinated effort is needed to address all areas of data acquisition in a systematic fashion. As stated in Chapter II, these areas include:

- Data collection request and approval procedures to reduce duplicate data collection
- Forms design and instruction writing guidelines to assure uniformity and clarity in data collection forms and definitions

- Document tracking procedures to provide an audit trail for locating stalled or lost source data

- Key verification techniques to control the accuracy of data entry

- Control and hash totals to assure complete and accurate processing of source data

(b) Data Element Dictionary/Directory

Our recommendation is to implement a passive data element dictionary/directory (DED/D) on a hierarchical basis. This allows the utilizaiton of the current DED/Ds in operation in EPA with the development of a centralized directory to be used to point to specific data elements in the detailed programmatic/system DED/D. Standards should be set for the development of additional DED/D. The current MIDSD effort to build a DED/D for selected systems using IDMS is the vehicle for not only setting standards but for eventually providing DED/D capabilities to the other programmatic areas. Finally, we recommend a detailed feasibility study for the implementation of an active DED/D in conjunction with the 1980 ADPE Procurement. This study should be conducted relatively soon so that the ADP specification for supporting an active DED/D can be included in the RFQ for the Procurement, if the study deems the system feasible.

(c) Feasibility Studies

The current procedures in the ADP Manual and those practiced by MIDSD appear to be quite effective. The proposed Mini-computer Review Group comprised of the Regional and Laboratory site ADP Managers can maintain effective control on standards for the Regions provided sufficient interaction with national level personnel is maintained. Our recommendation in this area is that the Agency ADP Oversight Committee endorse the review process by MIDSD and that the Committee members from each programmatic area personally see that the procedures are followed for their programmatic systems.

(d) Quality Assurance Program

We have addressed the problem of poor quality and lack of confidence in even some data that appears to be good. There are numerous efforts going on in Headquarters and the Regions to "clean up" the various data bases. These include the GICS and NEEDS systems in the Wastewater Treatment (WWT) program. Reconciliation and monitoring in the Financial Management System and numerous other efforts. What is needed is a coordinated effort supported by top management to first get all systems groups to initiate such system audit activities and second to assure that the programs do not focus just on "clearing up" the current

files but address and correct the causes that allowed bad data to get out in the files in the first place. As identified in Chapter II the reasons can usually be traced to one of the following:

- Lack of incentive on the part of those entering the data
- Confusion as to what data is required - content or form
- Lack of effective software edit capabilities.

The steps that can be taken to improve the quality of the data include:

- Inclusion of Effective Software in Systems -- edit and update procedures are the final line of defense against bad data being stored on files. They must be well conceived and, at times, imaginative.
- Standardization of Data Definitions -- standard definitions eliminate confusion on data to be collected.
- Revision of Forms and Procedures -- forms and procedures need to be easy to fill out and follow. Any ambiguity or unnecessary complexity should be eliminated.
- Introduction of New Incentives -- The data collector must be given an incentive to collect data. The best method is to offer reports or data access techniques that are of use to the collector. If this cannot be done then it is necessary to explore alternative means for collecting the data.
- Acquisition of Contractor Support -- It may be impossible to provide incentives to the current collectors for all or part of the collection and data entry process. If this occurs an alternative is to contract the work to private organizations. This is often done for data entry where incentive fees can provide the motive for accurate data entry. This method has been used in the NEEDS SURVEY, for the initial collection of the data.

Periodic audits should be performed and mechanisms should be built into the systems to identify problems. For example, a counter in the edit program for a specific error on a particular data element may help identify a definition or procedure that is ambiguous and needing revision. Consistency checks between similar items or the same data on two separate files can be performed. If there are significant discrepancies then a full audit, with the appropriate analysis and correction cycles would be mandated. These steps need to be formalized into policies and procedures that must be included in any revised set of Agency ADP/IS Policy and Procedures Manual.

(3) Policy and Procedures

The basic policies and procedures with regard to data management documents in the current ADP Manual indicate that the basis for control and an effective program are present, but are not currently implemented. There are many factors contributing to their not being implemented. These include:

- Decentralized management of data and systems
- Little high level mangement awareness of the need for data management
- Limited resources in MIDSD

We believe that the organization structure and tools recommended will provide the dynamic program structure and effective tools that will make the policies and procedures work. What is needed is a thorough review of the ADP manual to evaluate the appropriateness of the policies and procedures in the context of the other recommendations. Where changes are necessary they must be made. This can either be done by delegating this responsibility to the group establishing the new documentation standards or by establishing a separate task force. The latter approach would probably be preferred because of the combined workload. Finally, top level management commitment to the enforcement of these new policies and procedures must be obtained. The vehicle for this approval is the ADP Oversight Committee.

(4) Conclusion

The key to the successful implementation of this Data Management and Standardization Program is the combined support of top level management and the programmatic areas. This is normally a difficult request but we believe that it is possible because of the potential benefits to both. These potential benefits include:

- Facilitates sharing and re-utilization of data to ease the effects of potential resource reductions in some programs, and to reduce the reporting burden on the public and private sectors
- Provides a focal point for locating answers to Congressional, OMB, GAO, and FOI inquiries
- Provides Standards that Regions and States will be required to follow as programs are delegated
- Eases potential conversion efforts for the 1980 ADPE Procurement

- Provides an implementation program for the IRLG Common Codes Project.

The next chapter presents a plan for implementing the recommended program and provides cost estimates for the implementation.

Implementation Plan

IV. IMPLEMENTATION PLAN

IV. IMPLEMENTATION PLAN

Implementation of the proposed EPA Data Management and Standardization Program will require a nine step process which is graphically depicted in Exhibit IV-1. The first phase of this exhibit, Data Management and Standardization Program Plan, is shaded to indicate that it has been completed with the presentation of this report. Thus far the requirement for an EPA-wide Data Management and Standardization Program has been established; and program components have been defined. This information will serve as a starting point for the next phase of implementation, establishment of a data management task force.

There are two primary facets of the implementation plan:

- the implementation plan itself which includes activities, schedules, and control points; and
- supporting costs.

These two factors are further discussed in the sections below.

1. IMPLEMENTATION PLAN

To further present an expanded overview of the activities which are supportive to program implementation, this section has been organized into the following areas:

- Recommended Approach to Program Implementation
- Implementation Activties
- Discussion of Milestones and Control Points.

These aspects of the implementation plan are further discussed below.

(1) Recommended Approach to Program Implementation

Arthur Young & Company recommends a phased approach using phased program implementations of data management and standardization program concepts. Use of phased program implementations will:

- Spread the resource requirements for implementation over a longer period of time thereby balancing the use of personnel and funds

EPA Data Management and Standardization Program Life Cycle

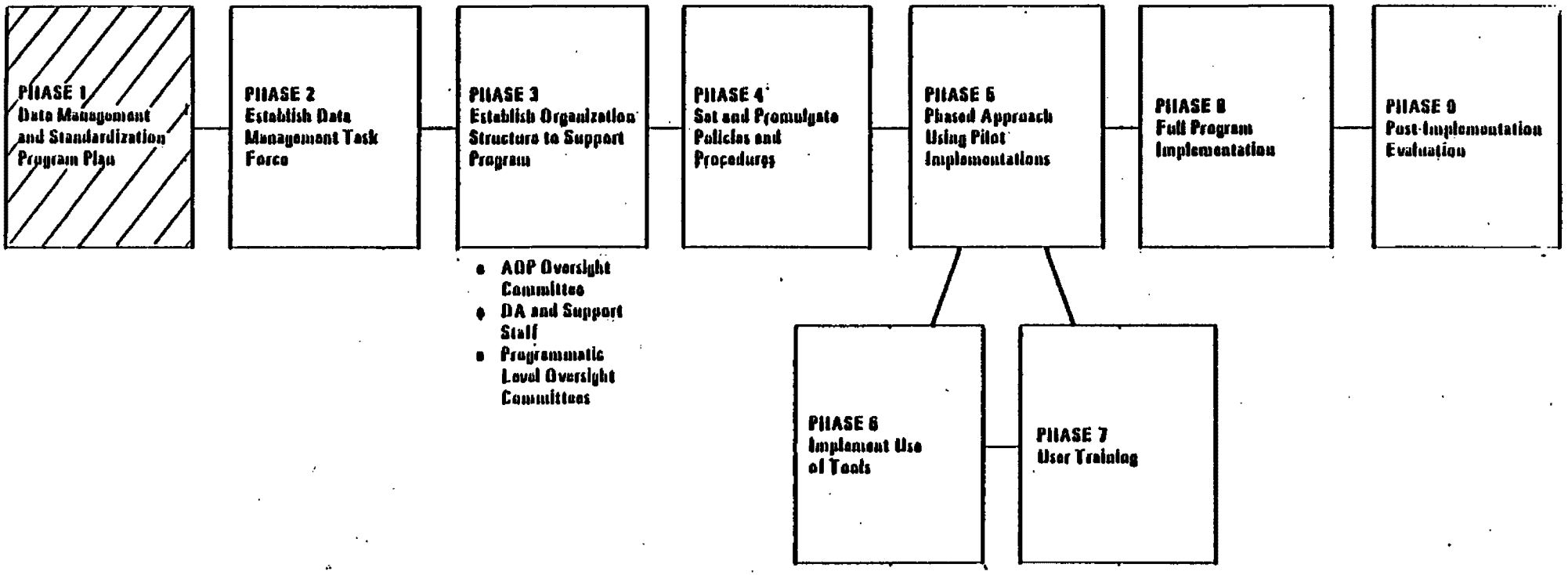


EXHIBIT IV-1

- Permit appropriate lead time for interfacing programmatic areas to prepare for program implementation

- Enable the proposed ADP Oversight Committee and supporting Data Administrator and staff to concentrate on the more detailed aspects of interfacing specific programmatic areas as program participants.

The activities which must be undertaken to establish the organization structure necessary to control phased program implementation and required implementation steps are further discussed in the section below.

(2) Implementation Activities

Implementation activities include those phases specified in Exhibit IV-1, these are the activities that follow the program plan to effect a full program implementation. The objectives and activities of the phases are discussed below:

(a) Establish Data Management Task Force

The first step EPA must take to develop a Data Management and Standardization Program is to establish a Data Management Task Force. The functions of this task force will include:

- Identification and nomination of members for the ADP Oversight Committee and supporting oversight committees throughout EPA

- Prepare draft procedures for day-to-day program operations

- Recommend the levels of administrative authority and responsibility for day-to-day program operations and the method for delegation throughout the organization.

The Data Management Task Force will provide the basis for formation of the required hierarchy of oversight committees. This group will also prepare the preliminary program procedures for ADP Oversight Committee evaluation. A final report should be produced to document all task force activities.

(b) Establish Organization Structure to Support the Program

Based on task force recommendations the ADP Oversight Committee and programmatic level oversight committees will be formed. The defined role of the ADP oversight committee should include:

- Establishment of data management program policies and procedures

- Specification and control of the use of data management tools

- Review and activate standards for:

  - System documentation

  - Data element definitions

  - Data element coding schemes

- Development, review and revision of long-range plans for data management and standardization program application

- Coordination of the data management activities of programmatic level oversight committees.

The defined role of the programmatic level oversight committees should include:

- Establishment of more detailed data management policy and procedures for their area which are consistent with the EPA-wide policies and procedures.

- Review and activate standards for:

  - System Documentation

  - Data Element Definitions

  - Data Element Coding Schemes

- Develop, review, and revise programmatic level Long-Range Plan for Data Management

- Prioritize a yearly plan for the development of new information systems and enhancement of current systems

- Review and endorse the information systems annual budget requests.

Both levels of the oversight committees will be required to monitor the adherence to their decisions through appropriate support staff.

The first task of these committees will be to agree on their working and meeting procedures and to identify the authority and responsibility which will be delegated to committee support staffs. The ADP Oversight Committee must also evaluate the skill levels for the Data Administrator and support staff positions and initiate activities to obtain the required personnel resources. Draft procedures for day-to-day program

administration must also be reviewed, modified, and finally adopted.

After the operational aspects of the ADP Oversight Committee and programmatic oversight committees are determined, a plan and schedule for phased program implementations must be developed and approved. A list of priorities to be considered during phased program implementations must also be developed and approved. Programmatic level oversight committees can contribute valuable support for this effort.

At the conclusion of this effort the EPA-wide ADP Oversight Committee, and supporting Data Administrator and staff should be organized and capable of functioning. Programmatic level oversight committees should be identified and committee recommendations and operations formally documented for activation prior to initiation of pilot program implementations in that area. All related components of the designated organization structures should be formally documented.

(c) Set Policies and Procedures

In this phase, the ADP Oversight Committee will formalize the data management and standardization policies and procedures which will govern program operations. These policies and procedures will be passed to the programmatic level oversight committees to serve as a basis for planning pilot program implementations, actual implementation, and ongoing program operations. The ADP Data Management and Standardization Program Policies and Procedural Manual being prepared by MIDSD should be finalized.

(d) Phased Approach Using Pilot Implementations

Based on the program implementation schedule approved in the previous phase by the ADP Oversight Committee, implementation of the data management program can be initiated. The programmatic level oversight committee, responsible for the new program participant area, should first be activated. The role of required committee support staff should be defined and staff composition determined. This role will be dependent on the implementation priorities established by the oversight committee and the level of authority and responsibility delegated to the oversight committee by the ADP oversight committee. If necessary, individuals will have to be transferred, hired and/or trained to meet the requirements. Once the staff is assembled work on installation of the tools for data management and standardization, and appropriate user training can begin. The plans for pilot program implementation should be formally documented for review by the ADP oversight committee and responsible programmatic oversight committee. This will be highly dependent on the current pilot projects in development. The phased approach comes into

play by adding other programmatic areas to the particular tool in a sequential rather than a one-shot implementation effort.

(e) Implement Use of Tools

The data management tools to be installed during pilot program implementations include:

- Application of the standards established by the ADP Oversight Committee for system development and documentation, data elements, and data acquisition techniques.
- Development of an Agency-wide data directory and individual program DED/D.
- Coordination procedures for feasibility studies for system design, system changes, and data standardization efforts
- Initiation of quality assurance audits which will identify the completeness and accuracy of system data, and analyze procedures for data capture and processing to isolate the causes of the errors.

Details of the composition of the specific tools are presented in Chapter III.

(f) User Training

An essential activity during phase of program implementation is the training of data management program participants. EPA personnel must be trained to accept program operations in a manner that will encourage active participation and promote the maintenance of a highly reliable program. Appropriate training materials related to EPA-wide data management and standardization concepts should be prepared by the Data Administrator support staff and approved by the ADP oversight committee. These training guides can then be distributed through programmatic level oversight committees to the committee support staff responsible for user training. This support staff should also prepare additional training materials related to the more detailed data management activities of their programmatic area. After approval by the responsible oversight committee, a combined training program can be initiated. User training manuals should also be produced for use as training guidelines and to document the training methodology employed.

(g) Full Program Implementation

At the conclusion of the final phased program implementation, all EPA organizational components should be participating in an active, organization-wide Data Management and Standardization

Program. Due to the large scale nature of this activity the ADP Oversight Committee may be required to reassess and revise original working and meeting procedures, and delegated authority and responsibility. In addition, this committee should begin to plan additional information management activities over a succeeding five year time period.

(h) Post-Implementation Evaluation

Post-implementation evaluation involves the determination of whether the fully implemented Data Management and Standardization Program meets EPA's needs and performs efficiently. This activity should be performed several months after the final pilot program implementation. This time lag is required so that EPA can form an opinion as to the program's effectiveness and efficiency.

A Post-Implementation Evaluation report should include an evaluation of program performance, operational costs, areas for improvements, and a determination of further enhancement areas.

(3) Milestones and Control Points

Each phase of the program implementation results in the production of a document which relates the activities performed during that phase. This document should be submitted to EPA management to facilitate review of activities to date, and adherence to schedules so that approval can be obtained for progression to the next phase of the implementation. Exhibit IV-2, on the following page, presents an approximate schedule for the implementation plan described in this chapter.

Throughout the program implementation considerations of costs, and availability of personnel resources may impact the implementation schedule presented in Exhibit IV-2. The evaluation of the entire implementation task at the completion of the activities of each phase provides milestones which will assist EPA management in their evaluation of potential impact on the program implementation effort and time schedule as a whole.

2. PROGRAM COST ESTIMATES

This section describes the methodology utilized by Arthur Young & Company to develop the cost estimates for the development and annual operations of the proposed EPA Data Management and Standardization Program. Program costs are discussed in terms of specific assumptions which provided the basis for cost algorithms; and a cost estimate matrix which summarizes the results of detailed cost computations and permits analysis from varying perspectives. The specific assumptions and cost estimates for the recommended program are further discussed below.

EPA Data Management and Standardization Program
Implementation Schedule

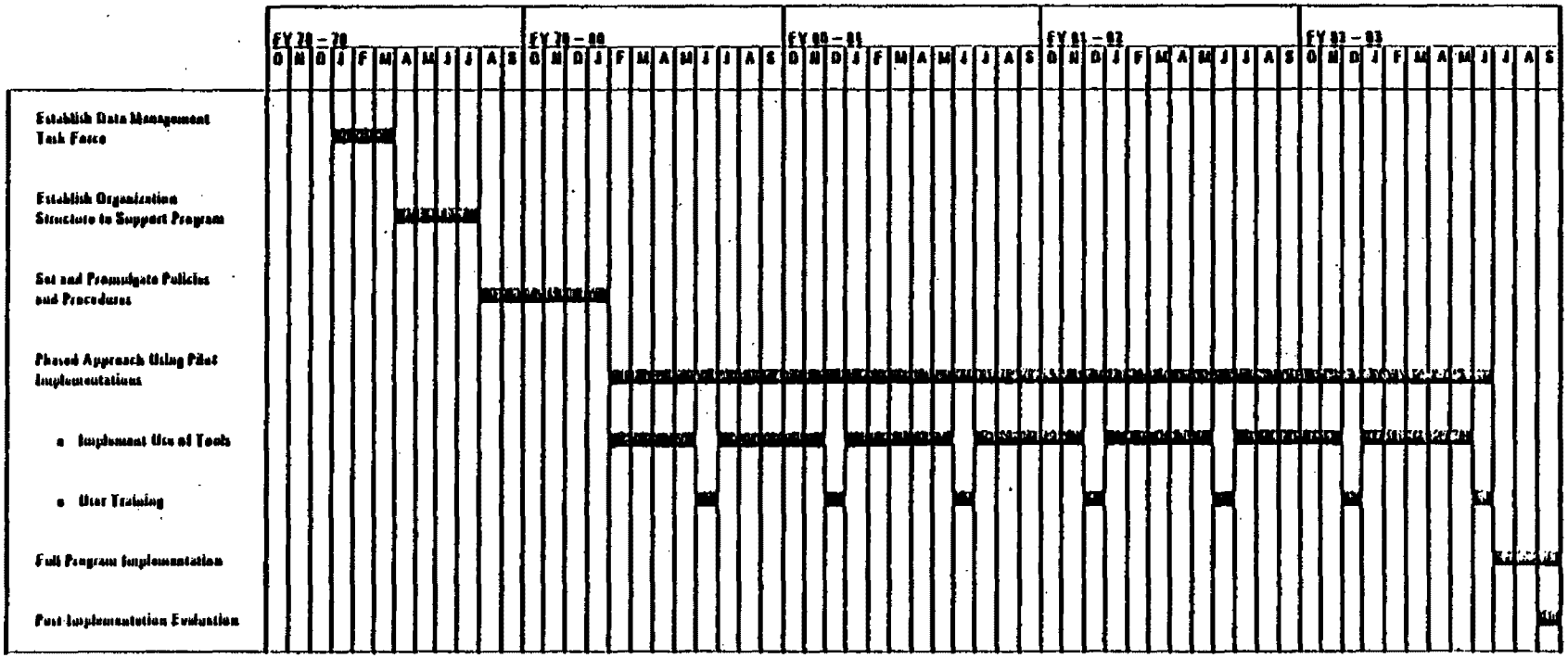


EXHIBIT IV-2

(1) Specific Assumptions

The rationale for the inclusion of specific cost assumptions are presented below:

- EPA Clerical Personnel - EPA clerical personnel will primarily be required to provide support services for the organizational component of the program structure. (ADP oversight committee, programmatic oversight committees, and data administrator and staff). This category does not include the costs for the preparation of documentation required to document each implementation phase. EPA clerical rates were costed at an average GS-7 Step 5 annual salary of $13,980. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- EPA Professional Personnel - EPA professional personnel will be required to actively participate in all phases of the program life cycle and annual operations. This category includes the EPA management personnel who will serve on the ADP oversight committee, programmatic oversight committees, or act as Data Administrator. Professional personnel will also attend user training sessions. Throughout annual program operations, professional personnel will be required for program administration and enforcement purposes. EPA professional personnel rates were costed at an average of GS-14 Step 5 annual salary of $34,850. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- EPA Systems Personnel - EPA systems personnel will support all phases of the program life cycle and annual operations. Primarily this category consists of programmatic level committee members and support staff, Data Administrator support staff, and Data Coordinators.

  On an ongoing operational basis, EPA systems personnel resources will support program operations related to the operational aspects of data management tools, participation in programmatic level oversight committees, and aspects of the Data Coordinator role which pertain to program activities or EPA passive data dictionary maintenance and use.

  The program-related costs for this type of personnel do not include the salaries for all systems personnel in EPA. The intent is to show only those costs which relate to utilization of systems personnel resources for data management and standardization program related activities. EPA systems personnel rates were costed at an average GS-12 Step 5 annual salary of $24,799. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- Documentation - At the conclusion of each phase of the program life cycle, various forms of documentation will be produced. This cost category includes expenses which would be incurred for the typing, graphic arts, and reproduction aspects of program documentation. Typing costs were based on 4 pages per hour at the EPA clerical rate of $7 per hour. Graphic arts preparation was based on 45 minutes per graphic at the EPA clerical rate of $7 per hour. Reproduction costs were based on $.05 per page, per copy produced.

- Other Assumptions - For the purpose of costing pilot program implementation and the resulting increases in annual operations costs, the following assumptions were made:

  - Approximately 15 pilot program implementations are required. Currently, there are approximately 22 EPA Deputy Assistant Administrators (DAA's) who could install programmatic level oversight committees to control data management activities in their areas. Each of these DAA's may not, however, have sufficient systems activity to warrant program implementation. In addition, the systems activity for certain areas may fall within the scope of other programmatic level oversight committees. Consideration should also be given to the fact that some DAA's may wish to jointly participate in the program through one common oversight committee due to responsibilities for highly related systems. It was therefore determined that at least 11 DAA's would implement their own program, while the remaining 11 would jointly share in 4 other oversight committees.

  - Pilot program implementation would be undertaken based on the following schedule: 0 in fiscal year 78-79; three in fiscal year 79-80; and four each in fiscal years 80-81, 81-82, and 82-83.

(2) Cost Matrix

To present the cost estimates for the implementation and annual operations of an EPA Data Management and Standardization Program, cost estimate matrices were utilized. Since the recommended implementation plan for this program includes a phased approach consisting of pilot program implementations over a five-year period, cost estimate matrices have been prepared for each year of the five-year implementation plan. These annual cost estimate matrices are shown as Exhibits IV-3 through IV-7. In addition, total costs for the five-year implementation plan are shown on Exhibit IV-8. Exhibit IV-9 shows the total annual operations costs for the program after full implementation, for an additional five-year period. A 7% cost escalation factor was added to all cost estimates for each successive year beyond fiscal

- Documentation - At the conclusion of each phase of the program life cycle, various forms of documentation will be produced. This cost category includes expenses which would be incurred for the typing, graphic arts, and reproduction aspects of program documentation. Typing costs were based on 4 pages per hour at the EPA clerical rate of $7 per hour. Graphic arts preparation was based on 45 minutes per graphic at the EPA clerical rate of $7 per hour. Reproduction costs were based on $.05 per page, per copy produced.

- Other Assumptions - For the purpose of costing pilot program implementation and the resulting increases in annual operations costs, the following assumptions were made:

  - Approximately 15 pilot program implementations are required. Currently, there are approximately 22 EPA Deputy Assistant Administrators (DAA's) who could install programmatic level oversight committees to control data management activities in their areas. Each of these DAA's may not, however, have sufficient systems activity to warrant program implementation. In addition, the systems activity for certain areas may fall within the scope of other programmatic level oversight committees. Consideration should also be given to the fact that some DAA's may wish to jointly participate in the program through one common oversight committee due to responsibilities for highly related systems. It was therefore determined that at least 11 DAA's would implement their own program, while the remaining 11 would jointly share in 4 other oversight committees.

  - Pilot program implementation would be undertaken based on the following schedule: 0 in fiscal year 78-79; three in fiscal year 79-80; and four each in fiscal years 80-81, 81-82, and 82-83.

(2) Cost Matrix

To present the cost estimates for the implementation and annual operations of an EPA Data Management and Standardization Program, cost estimate matrices were utilized. Since the recommended implementation plan for this program includes a phased approach consisting of pilot program implementations over a five-year period, cost estimate matrices have been prepared for each year of the five-year implementation plan. These annual cost estimate matrices are shown as Exhibits IV-3 through IV-7. In addition, total costs for the five-year implementation plan are shown on Exhibit IV-8. Exhibit IV-9 shows the total annual operations costs for the program after full implementation, for an additional five-year period. A 7% cost escalation factor was added to all cost estimates for each successive year beyond fiscal

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 78-79

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	$ 1,400	$ 49,000	$ 9,200	$ 2,000	$ 2,000	$ 59,600	$ 61,600
	ESTABLISH ORGANIZATION STRUCTURE	$ 4,000	$ 32,200	$127,000	$ 1,400	$ 1,400	$ 163,200	$ 164,600
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 1,400	$ 20,100	$ 85,900	$ 1,400	$ 1,400	$ 107,400	$ 108,800
TOTAL DEVELOPMENT COSTS		$ 6,800	$ 101,300	$222,100	$ 4,800	$ 4,800	$ 330,200	$ 335,000
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	---	---	---	---	---	---	---
	USER TRAINING	---	---	---	---	---	---	---
TOTAL PROGRAM IMPLEMENTATION COSTS		---	---	---	---	---	---	---
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 6,800	$ 101,300	$222,100	$ 4,800	$ 4,800	$ 330,200	$ 335,000
ANNUAL PROGRAM OPERATIONS COSTS		$28,000	$ 72,400	$ 74,400	$ 2,200	$ 2,200	$ 174,800	$ 177,000

IV. IMPLEMENTATION PLAN

Implementation Plan

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 79-80

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—	—	
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—	—	—	—	
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 3,000	$ 43,000	$ 183,900	$ 2,600	$ 2,600	$ 229,900	$ 232,500
TOTAL DEVELOPMENT COSTS		$ 3,000	$ 43,000	$ 183,900	$ 2,600	$ 2,600	$ 229,900	$ 232,500
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 21,600	$ 53,800	$ 153,000	$ 4,900	$ 4,900	$ 228,400	$ 233,300
	USER TRAINING	$ 400	$ 4,300	$ 9,200	$ 3,300	$ 3,300	$ 13,900	$ 17,200
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 22,000	$ 58,100	$ 162,200	$ 8,200	$ 8,200	$ 242,300	$ 250,500
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 25,000	$ 101,100	$ 346,100	$ 10,800	$ 10,800	$ 472,200	$ 483,000
ANNUAL PROGRAM OPERATIONS COSTS		$ 85,300	$ 146,500	$ 177,900	$ 9,900	$ 9,900	$ 409,700	$ 419,600

EXHIBIT IV-4

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 80-81

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—			—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—	—	—	—	—		
TOTAL DEVELOPMENT COSTS		—	—	—	—	—		
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 30,800	$ 76,700	$ 218,300	$ 7,000	$ 7,000	$ 325,800	$ 332,800
	USER TRAINING	$ 500	$ 6,200	$ 13,100	$ 4,700	$ 4,700	$ 19,800	$ 24,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 31,300	$ 82,900	$ 231,400	$ 11,700	$ 11,700	$ 345,600	$ 357,300
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 31,300	$ 82,900	$ 231,400	$ 11,700	$ 11,700	$ 345,600	$ 357,300
ANNUAL PROGRAM OPERATIONS COSTS		$ 229,600	$ 329,100	$ 435,700	$ 29,700	$ 29,700	$ 994,400	$1,024,100

EXHIBIT IV-5

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 81-82

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPOR-TUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—		—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—		—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—	—	—		—		
TOTAL DEVELOPMENT COSTS		—	—	—		—		
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 33,000	$ 82,100	$ 233,600	$ 7,500	$ 7,500	$ 348,700	$ 356,200
	USER TRAINING	$ 600	$ 6,700	$ 14,100	$ 5,100	$ 5,100	$ 21,400	$ 26,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 33,600	$ 88,800	$ 247,700	$ 12,600	$ 12,600	$ 370,100	$ 382,700
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 33,600	$ 88,800	$ 247,700	$ 12,600	$ 12,600	$ 370,100	$ 382,700
ANNUAL PROGRAM OPERATIONS COSTS		$ 414,000	$ 562,900	$ 766,200	$ 54,000	$ 54,800	$1,743,900	$1,798,700

EXHIBIT IV-6

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 82-83

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION		TOTALS	
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPOR-TUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—	—	—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—	—	—	—	—		
TOTAL DEVELOPMENT COSTS		—	—	—				
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 35,400	$ 87,900	$ 250,000	$ 8,100	$ 8,100	$ 373,300	$ 381,400
	USER TRAINING	$ 700	$ 7,200	$ 15,100	$ 5,500	$ 5,500	$ 23,000	$ 28,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 36,100	$ 95,100	$ 265,100	$ 13,600	$ 13,600	$ 396,300	$ 409,900
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 36,100	$ 95,100	$ 265,100	$ 13,600	$ 13,600	$ 396,300	$ 409,900
ANNUAL PROGRAM OPERATIONS COSTS		$ 624,000	$ 827,800	$1,140,900	$ 83,300	$ 83,300	$2,593,500	$2,676,800

EXHIBIT IV-7

EPA Data Management and Standardization Program

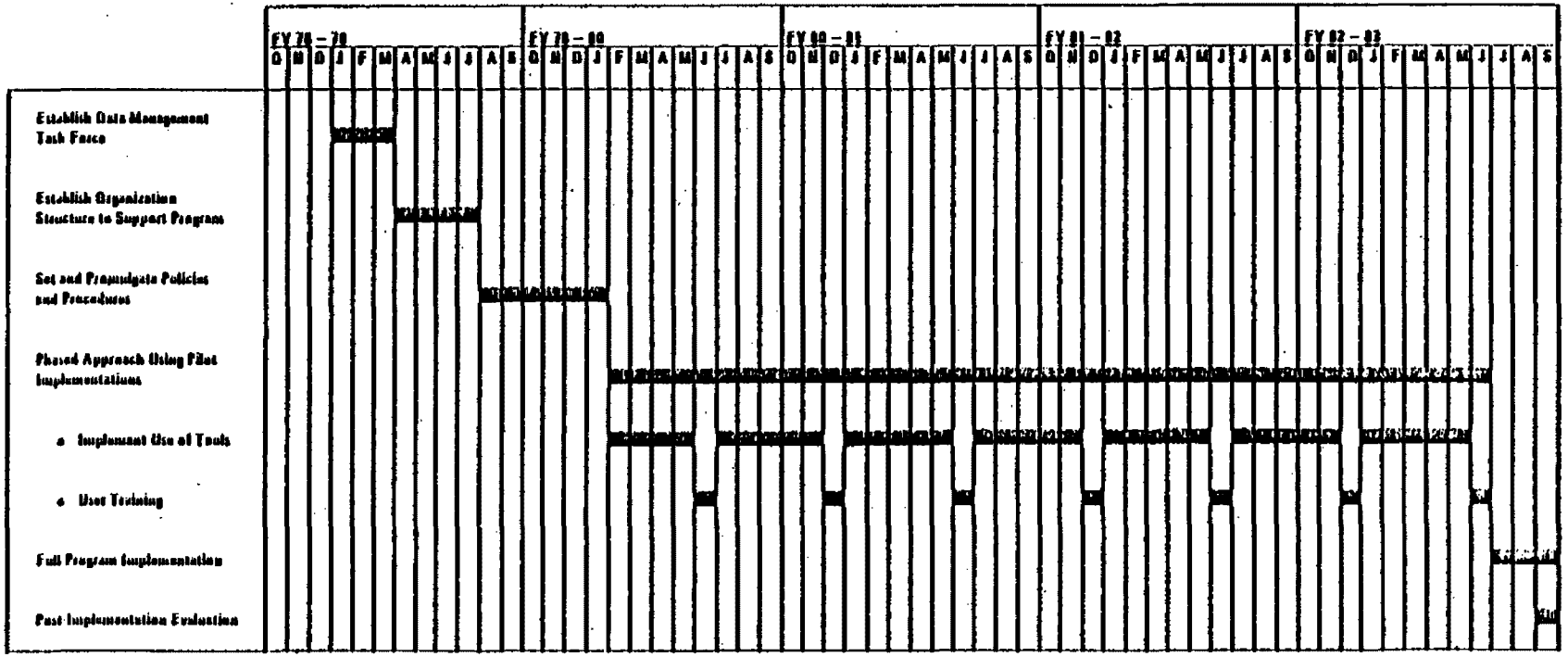
COST ESTIMATE MATRIX

COST SUMMARY - FIVE-YEAR PLAN

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	$ 1,400	$ 49,000	$ 9,200	$ 2,000	$ 2,000	$ 59,600	$ 61,600
	ESTABLISH ORGANIZATION STRUCTURE	$ 4,000	$ 32,200	$ 127,000	$ 1,400	$ 1,400	$ 163,200	$ 164,600
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 4,400	$ 63,100	$ 269,800	$ 4,000	$ 4,000	$ 337,300	$ 341,300
TOTAL DEVELOPMENT COSTS		$ 9,800	$ 144,300	$ 406,000	$ 7,400	$ 7,400	$ 560,100	$ 567,500
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$120,000	$ 300,500	$ 854,900	$ 27,500	$ 27,500	$1,276,200	$1,303,700
	USER TRAINING	$ 2,200	$ 24,400	$ 51,500	$ 18,600	$ 18,600	$ 78,100	$ 96,700
TOTAL PROGRAM IMPLEMENTATION COSTS		$123,000	$ 324,900	$ 906,400	$ 46,100	$ 46,100	$1,354,300	$1,400,400
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$132,000	$ 469,200	$1,312,400	$ 53,500	$ 53,500	$1,914,400	$1,967,900
ANNUAL PROGRAM OPERATIONS COSTS		$1,382,500	$1,938,700	$2,595,100	$ 179,900	$ 179,900	$5,916,300	$6,096,200

EXHIBIT IV-8

EPA Data Management and Standardization Program
Implementation Schedule



(1) Specific Assumptions

The rationale for the inclusion of specific cost assumptions are presented below:

- EPA Clerical Personnel - EPA clerical personnel will primarily be required to provide support services for the organizational component of the program structure. (ADP oversight committee, programmatic oversight committees, and data administrator and staff). This category does not include the costs for the preparation of documentation required to document each implementation phase. EPA clerical rates were costed at an average GS-7 Step 5 annual salary of $13,980. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- EPA Professional Personnel - EPA professional personnel will be required to actively participate in all phases of the program life cycle and annual operations. This category includes the EPA management personnel who will serve on the ADP oversight committee, programmatic oversight committees, or act as Data Administrator. Professional personnel will also attend user training sessions. Throughout annual program operations, professional personnel will be required for program administration and enforcement purposes. EPA professional personnel rates were costed at an average of GS-14 Step 5 annual salary of $34,850. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- EPA Systems Personnel - EPA systems personnel will support all phases of the program life cycle and annual operations. Primarily this category consists of programmatic level committee members and support staff, Data Administrator support staff, and Data Coordinators.

  On an ongoing operational basis, EPA systems personnel resources will support program operations related to the operational aspects of data management tools, participation in programmatic level oversight committees, and aspects of the Data Coordinator role which pertain to program activities or EPA passive data dictionary maintenance and use.

  The program-related costs for this type of personnel do not include the salaries for all systems personnel in EPA. The intent is to show only those costs which relate to utilization of systems personnel resources for data management and standardization program related activities. EPA systems personnel rates were costed at an average GS-12 Step 5 annual salary of $24,799. Hourly rates were calculated utilizing a 260 day work year comprised of 2080 hours.

- Documentation - At the conclusion of each phase of the program life cycle, various forms of documentation will be produced. This cost category includes expenses which would be incurred for the typing, graphic arts, and reproduction aspects of program documentation. Typing costs were based on 4 pages per hour at the EPA clerical rate of $7 per hour. Graphic arts preparation was based on 45 minutes per graphic at the EPA clerical rate of $7 per hour. Reproduction costs were based on $.05 per page, per copy produced.

- Other Assumptions - For the purpose of costing pilot program implementation and the resulting increases in annual operations costs, the following assumptions were made:

  - Approximately 15 pilot program implementations are required. Currently, there are approximately 22 EPA Deputy Assistant Administrators (DAA's) who could install programmatic level oversight committees to control data management activities in their areas. Each of these DAA's may not, however, have sufficient systems activity to warrant program implementation. In addition, the systems activity for certain areas may fall within the scope of other programmatic level oversight committees. Consideration should also be given to the fact that some DAA's may wish to jointly participate in the program through one common oversight committee due to responsibilities for highly related systems. It was therefore determined that at least 11 DAA's would implement their own program, while the remaining 11 would jointly share in 4 other oversight committees.

  - Pilot program implementation would be undertaken based on the following schedule: 0 in fiscal year 78-79; three in fiscal year 79-80; and four each in fiscal years 80-81, 81-82, and 82-83.

(2) Cost Matrix

To present the cost estimates for the implementation and annual operations of an EPA Data Management and Standardization Program, cost estimate matrices were utilized. Since the recommended implementation plan for this program includes a phased approach consisting of pilot program implementations over a five-year period, cost estimate matrices have been prepared for each year of the five-year implementation plan. These annual cost estimate matrices are shown as Exhibits IV-3 through IV-7. In addition, total costs for the five-year implementation plan are shown on Exhibit IV-8. Exhibit IV-9 shows the total annual operations costs for the program after full implementation, for an additional five-year period. A 7% cost escalation factor was added to all cost estimates for each successive year beyond fiscal

Documentation - At the conclusion of each phase of the program life cycle, various forms of documentation will be produced. This cost category includes expenses which would be incurred for the typing, graphic arts, and reproduction aspects of program documentation. Typing costs were based on 4 pages per hour at the EPA clerical rate of $7 per hour. Graphic arts preparation was based on 45 minutes per graphic at the EPA clerical rate of $7 per hour. Reproduction costs were based on $.05 per page, per copy produced.

Other Assumptions - For the purpose of costing pilot program implementation and the resulting increases in annual operations costs, the following assumptions were made:

- Approximately 15 pilot program implementations are required. Currently, there are approximately 22 EPA Deputy Assistant Administrators (DAA's) who could install programmatic level oversight committees to control data management activities in their areas. Each of these DAA's may not, however, have sufficient systems activity to warrant program implementation. In addition, the systems activity for certain areas may fall within the scope of other programmatic level oversight committees. Consideration should also be given to the fact that some DAA's may wish to jointly participate in the program through one common oversight committee due to responsibilities for highly related systems. It was therefore determined that at least 11 DAA's would implement their own program, while the remaining 11 would jointly share in 4 other oversight committees.

- Pilot program implementation would be undertaken based on the following schedule: 0 in fiscal year 78-79; three in fiscal year 79-80; and four each in fiscal years 80-81, 81-82, and 82-83.

(2) Cost Matrix

To present the cost estimates for the implementation and annual operations of an EPA Data Management and Standardization Program, cost estimate matrices were utilized. Since the recommended implementation plan for this program includes a phased approach consisting of pilot program implementations over a five-year period, cost estimate matrices have been prepared for each year of the five-year implementation plan. These annual cost estimate matrices are shown as Exhibits IV-3 through IV-7. In addition, total costs for the five-year implementation plan are shown on Exhibit IV-8. Exhibit IV-9 shows the total annual operations costs for the program after full implementation, for an additional five-year period. A 7% cost escalation factor was added to all cost estimates for each successive year beyond fiscal

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 78-79

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	$ 1,400	$ 49,000	$ 9,200	$ 2,000	$ 2,000	$ 59,600	$ 61,600
	ESTABLISH ORGANIZATION STRUCTURE	$ 4,000	$ 32,200	$127,000	$ 1,400	$ 1,400	$ 163,200	$ 164,600
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 1,400	$ 20,100	$ 85,900	$ 1,400	$ 1,400	$ 107,400	$ 108,800
TOTAL DEVELOPMENT COSTS		$ 6,800	$ 101,300	$222,100	$ 4,800	$ 4,800	$ 330,200	$ 335,000
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	---	---	---	---	---	---	---
	USER TRAINING	---	---	---	---	---	---	---
TOTAL PROGRAM IMPLEMENTATION COSTS		---	---	---	---	---	---	---
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 6,800	$ 101,300	$222,100	$ 4,800	$ 4,800	$ 330,200	$ 335,000
ANNUAL PROGRAM OPERATIONS COSTS		$28,000	$ 72,400	$ 74,400	$ 2,200	$ 2,200	$ 174,800	$ 177,000

EXHIBIT IV-3

IV. IMPLEMENTATION PLAN

Implementation Plan

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 79-80

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPOR-TUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—	—	
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—	—	—	—	
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 3,000	$ 43,000	$ 183,900	$ 2,600	$ 2,600	$ 229,900	$ 232,500
TOTAL DEVELOPMENT COSTS		$ 3,000	$ 43,000	$ 183,900	$ 2,600	$ 2,600	$ 229,900	$ 232,500
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 21,600	$ 53,800	$ 153,000	$ 4,900	$ 4,900	$ 220,400	$ 233,300
	USER TRAINING	$ 400	$ 4,300	$ 9,200	$ 3,300	$ 3,300	$ 13,900	$ 17,200
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 22,000	$ 58,100	$ 162,200	$ 8,200	$ 8,200	$ 242,300	$ 250,500
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 25,000	$ 101,100	$ 346,100	$ 10,800	$ 10,800	$ 472,200	$ 483,000
ANNUAL PROGRAM OPERATIONS COSTS		$ 85,300	$ 146,500	$ 177,900	$ 9,900	$ 9,900	$ 409,700	$ 419,600

EXHIBIT IV-4

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 80-81

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION		TOTALS	
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—			—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—	—	—	—	—		
TOTAL DEVELOPMENT COSTS		—	—	—	—	—		
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 30,800	$ 76,700	$ 218,300	$ 7,000	$ 7,000	$ 325,800	$ 332,800
	USER TRAINING	$ 500	$ 6,200	$ 13,100	$ 4,700	$ 4,700	$ 19,800	$ 24,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 31,300	$ 82,900	$ 231,400	$ 11,700	$ 11,700	$ 345,600	$ 357,300
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 31,300	$ 82,900	$ 231,400	$ 11,700	$ 11,700	$ 345,600	$ 357,300
ANNUAL PROGRAM OPERATIONS COSTS		$ 229,600	$ 329,100	$ 435,700	$ 29,700	$ 29,700	$ 994,400	$1,024,100

EXHIBIT IV-5

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 81-82

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION		TOTALS	
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—	—	—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—	—	—	—	—		
TOTAL DEVELOPMENT COSTS		—	—	—	—	—		
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 33,000	$ 82,100	$ 233,600	$ 7,500	$ 7,500	$ 348,700	$ 356,200
	USER TRAINING	$ 600	$ 6,700	$ 14,100	$ 5,100	$ 5,100	$ 21,400	$ 26,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 33,600	$ 88,800	$ 247,700	$ 12,600	$ 12,600	$ 370,100	$ 382,700
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 33,600	$ 88,800	$ 247,700	$ 12,600	$ 12,600	$ 370,100	$ 382,700
ANNUAL PROGRAM OPERATIONS COSTS		$ 414,000	$ 562,900	$ 766,200	$ 54,000	$ 54,800	$1,743,900	$1,798,700

EXHIBIT IV-6

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

FISCAL YEAR 82-83

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	—	—	—	—	—		
	ESTABLISH ORGANIZATION STRUCTURE	—	—	—	—	—		
	SET AND PROMULGATE POLICIES AND PROCEDURES	—		—	—	—		
TOTAL DEVELOPMENT COSTS		—		—				
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$ 35,400	$ 87,900	$ 250,000	$ 8,100	$ 8,100	$ 373,300	$ 381,400
	USER TRAINING	$ 700	$ 7,200	$ 15,100	$ 5,500	$ 5,500	$ 23,000	$ 28,500
TOTAL PROGRAM IMPLEMENTATION COSTS		$ 36,100	$ 95,100	$ 265,100	$ 13,600	$ 13,600	$ 396,300	$ 409,900
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$ 36,100	$ 95,100	$ 265,100	$ 13,600	$ 13,600	$ 396,300	$ 409,900
ANNUAL PROGRAM OPERATIONS COSTS		$ 624,000	$ 827,800	$1,140,900	$ 83,300	$ 83,300	$2,593,500	$2,676,800

EXHIBIT IV-7

EPA Data Management and Standardization Program

COST ESTIMATE MATRIX

COST SUMMARY - FIVE-YEAR PLAN

LIFE CYCLE PHASE	COST ELEMENTS	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
		EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPORTUNITY COST	OVERALL TOTAL
DATA MANAGEMENT PROGRAM DEVELOPMENT COSTS	ESTABLISH DATA MANAGEMENT TASK FORCE	$ 1,400	$ 49,000	$ 9,200	$ 2,000	$ 2,000	$ 59,600	$ 61,600
	ESTABLISH ORGANIZATION STRUCTURE	$ 4,000	$ 32,200	$ 127,000	$ 1,400	$ 1,400	$ 163,200	$ 164,600
	SET AND PROMULGATE POLICIES AND PROCEDURES	$ 4,400	$ 63,100	$ 269,800	$ 4,000	$ 4,000	$ 337,300	$ 341,300
TOTAL DEVELOPMENT COSTS		$ 9,800	$ 144,300	$ 406,000	$ 7,400	$ 7,400	$ 560,100	$ 567,500
PROGRAM IMPLEMENTATION COSTS	IMPLEMENT USE OF TOOLS	$120,000	$ 300,500	$ 854,900	$ 27,500	$ 27,500	$1,276,200	$1,303,700
	USER TRAINING	$ 2,200	$ 24,400	$ 51,500	$ 18,600	$ 18,600	$ 78,100	$ 96,700
TOTAL PROGRAM IMPLEMENTATION COSTS		$123,000	$ 324,900	$ 906,400	$ 46,100	$ 46,100	$1,354,300	$1,400,400
TOTAL PROGRAM DEVELOPMENT AND IMPLEMENTATION COSTS		$132,800	$ 469,200	$1,312,400	$ 53,500	$ 53,500	$1,914,400	$1,967,900
ANNUAL PROGRAM OPERATIONS COSTS		$1,302,500	$1,938,700	$2,595,100	$ 179,900	$ 179,900	$5,916,300	$6,096,200

EXHIBIT IV-8

EPA Data Management and Standardization Program

FIVE-YEAR COST AFTER FULL IMPLEMENTATION

FISCAL YEAR	PERSONNEL RESOURCES			DOCUMENTATION	TOTALS		
	EPA CLERICAL	EPA PROFESSIONAL	EPA SYSTEMS		OUT-OF POCKET EXPENSES	OPPOR-TUNITY COST	OVERALL TOTAL
83-84	$ 765,300	$ 1,006,300	$ 1,392,400	$ 102,300	$ 102,300	$3,164,000	$3,266,300
84-85	$ 818,900	$ 1,076,800	$ 1,489,900	$ 109,500	$ 109,500	$3,385,600	$3,495,100
85-86	$ 876,200	$ 1,152,100	$ 1,594,200	$ 117,100	$ 117,100	$3,622,500	$3,739,600
86-87	$ 937,500	$ 1,232,800	$ 1,705,800	$ 125,300	$ 125,300	$3,876,100	$4,001,400
87-88	$ 1,003,100	$ 1,319,100	$ 1,825,200	$ 134,100	$ 134,100	$4,147,400	$4,281,500
TOTAL	$ 4,401,000	$ 5,787,100	$ 8,007,500	$ 588,300	$ 588,300	$18,195,600	$18,783,900

EXHIBIT IV-9

years 78-79 to account for anticipated increases due to inflation. This percentage was based on trends in the consumer price index.

The cost estimate matrices cross tabulate the cost elements of the program implementation and operations life cycles. The component parts of each life cycle phase include sumarized costs which are based on algorithms utilized to develop costs for the types of resources for which specific assumptions were made. These resources include the following:

- Personnel Resources
  - EPA Clerical
  - EPA Professional
  - EPA Systems
- Documentation

Totals are expressed in terms of out-of-pocket expenses, opportunity cost and overall total. These components of the final totals are further explained below.

- Out-of-Pocket Expenses - Out-of-Pocket expenses represent those expenses which EPA will incur through contractual services or through direct purchases. These expenses must be included in the EPA budget and necessitate additional expenditures over current funds. Only the documentation produced during program development, implementation, and annual operations was categorized as an out-of-pocket expense.

- Opportunity Costs - Opportunity costs are the costs of personnel resources which must be redirected for program development, implementation, and annual operations efforts. Opportunity costs must be viewed as a loss or restriction of existing resources. For example, the opportunity cost of staff time represents the value of their salaries for the time which they will be required to devote to program activities. It is important to note that an opportunity cost is not necessarily an additional dollar expenditure with a direct budget impact, but is an analytical technique for evaluation of true program cost.

The costs for the program are expressed in terms of data management program development, program implementation, and annual operations. The first two cost factors consist of one-time expenditures, while the third factor constitutes ongoing expenses. Cost elements which comprise the development, implementation, and operations portions of the program life cycle cross-tabulate each of the resources previously discussed. All

costs have been rounded to the next hundred to simplify presentation.

costs have been rounded to the next hundred to simplify presentation.

Appendicies

APPENDIX A

DATA DICTIONARY

Appendix A

TABLE OF CONTENTS

INTRODUCTION

The following dictionary is a compilation of metadata for selected data categories from a representative sample of EPA systems. (The term metadata means information about the data itself.) The dictionary was developed to assist in determining the need for a data management and standardization program in EPA. The dictionary will be used to highlight problems with definitions, coding schemes and data elements within categories. For example, there currently exist data elements with the same name that have different definitions as well as different coding schemes. The dictionary was developed by reviewing system documentation and transferring information about selected data elements onto data collection sheets (identified by the column "sheet" in the dictionary). The information shown in this report is a subset of the information collected on the data collection sheets organized by data categories. The five categories of data elements are:

I - Facility Identifier

II - Sample Station Site

III - Facility Geographic Locations

IV - Parameter Unit Identifier

V - Quality Assurance Codes

The 15 Representative Systems included in this analysis are:

AEROS - National Emission Data System (NEDS)
AEROS - Storage and Retrieval of Aeromatic Data (SAROAD)
Compliance Data System (CDS)
Energy Data System (EDS)
Establishment Registration Support System (ERSS)
Financial Management System (FMS)
Grants Information & Control System (GICS)
Model States Information System (MSIS)
Estimate of Municipal Wastewater Treatment Requirements (NEEDS)
Pesticides Enforcement Monitoring System (PEMS)
Permit Compliance System (PCS)
Pesticides Product Information System (PPIS)
Regional Air Monitoring System (RAMS)
Storage and Retrieval of Water Quality Data (STORET)
Chemical Substance Inventory Report (TSCAS)

The seven metadata elements presented are:

System - Mnemonic of the systems in which the element is found

Name - Data element name that best describes the element as it is used in the representative EPA systems

Picture - Space allocated for the data element in the data base

Code - A key to the specific code used for the data element. The code sheets are presented in Appendix B

Sheet - A key to the data collection sheet, in the working papers that contains the detailed information about the specific data item

Source - The organization or form from which the information is obtained

Definition - The data element definition contained in the system documentation

DATA CATEGORY Facility Identifiers
DATA ELEMENT TYPE Authority/Company Name

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
EDS	Company Name	X(30)		ED-26	FPC 67/NEDS	
ERSS	Company Name	X(50)		ER-1	Registration Division	Complete name of the home office or headquarters of the company submitting the application.
MSIS	Plant Owner Name	X(30)		MS-4	State	Represents the name of the offical plant owner of a Public Water System
NEEDS	Authority Name	X(20)		NE-3	EPA-1	The official name of the authority will be the name which is used to legally identify it. Any unit of a State, county, or city government or any other non-Federal unit of government which is responsible for the collection and/or treatment of municipal wastewater.
PPIS	Company Name	X(40)		PS-2		Name of company.
TOXICS	Corporation	X(29)		TS-1	Chemical Substances Inventory Report	Name of the domestic corporation of which the plant site is a part of.

DATA CATEGORY Facility Identifiers
DATA ELEMENT TYPE Authority/Company Code

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
EDS	Company Code	X(6)		ED-3	FPC 67	
ERSS	Company Number	9(5)		ER-2	PARC/Application form	Company registration number (assigned by EPA)
FMS	Employer Identification No.	X(11)	C1	FM-2		Identify organizations who are recipients of EPA Grants. This code will be used in lieu of geographic location. This is the code for all grant programs except construction and fellowships (Above 96% exceptions).
NEEDS	Authority/Facility No.	X(9)	F2	NE-1	EPA 1	Unique identifier that consists of authority and facility number.
PPIS	EPA Company ID	X(6)		PS-8		Used to identify company name.
PPIS	Distributor No.	X(6)		PS-8		Same as the EPA company number component of the EPA registration number.
TOXICS	Dun & Bradstreet No.	X(11)		TS-8	Chemical Substances Inventory Report	Lists the corporation Dun Bradstreet registration number.
TOXICS	Manufactures ID	X(7)		TS-20		Unique indentifier for name and address of reporting person, site corporation or trade association.

DATA CATEGORY Facility Identifiers
DATA ELEMENT TYPE Facility Name

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Plant Name and Address	X(40)		AE-1		Plant Name and Address
CDS	Source Name	X(20)		CD-10	Region	Alphanumeric facility identifier
EDS	Plant Name	X(25)		ED-25	FPC 67/NEDS	
ERSS	Establishment Name	X(40)		ER-12	Establishment Registration Form	The name of all production sites whose production is controlled solely by the company.
* MSIS	Public Water System	X(30)		MS-23	State	The official name of the public water system.
NEEDS	Facility Name			NE-2	EPA-1	The name most frequently used by the authority to identify that facility.
PEMS	Establishment Name	X(40)		PM-3	Region	
PEMS	SMUD Name	X(30)		PM-6	Sample Collection Report	Shipper, Manufacturer, User, Distributor Number
PEMS	Port/Establishment Name	X(30)			Region	
PCS	Facility Name	X(60)		PR-2	NPDES Application Form	The official or legal name used to distinguish the facility from similar entities in the same geographical area.
RAMS	Plant	X(25)		RA-2		
TOXICS	Plant Site Name	X(30)		TS-2	Chemical Substance Inventory Report	Name of the plant where chemical substances are manufactured or processed.

* EPA has funded and maintained the Model States Information System (MSIS) but it is used by states only and not EPA. On an annual basis states can use MSIS to generate data which must be submitted to EPA for inclusion in the Federal Recording and Data System (FRDS). Where states won't take primacy for MSIS control, the EPA Regions act as surrogate states and enforce regulatory provisions which require collection of data for reporting in this system.

DATA CATEGORY Facility Identifiers
DATA ELEMENT TYPE Facility ID Number

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Plant Identification	X(4)	F5	AE-2		Plant identification number assigned by the county. The number can be alphabetic or numeric.
CDS	State Registration No.	X(15)		CD-14		Identifier used by the state to identify a facility.
CDS	Source Number	9(5)		CD-6	Region/NEDS	5-digit numeric facility identifier.
CDS	NEDS Cross Reference No.	X(4)	F5	CD-15	NEDS	NEDS facility identifier
EDS	Entity Identification	X(12)	F5,F6	ED-1	FPC 67/NEDS/SAROAD	FPC Plant Code, NEDS Code or SAROAD code.
EDS	NEDS Identification Code	X(10)	F5	ED-2	NEDS	NEDS Identification Code
EDS	Plant Code	X(4)	F6	ED-4	FPC 67	FPC Plant Code
ERSS	Establishment Sequence No.	9(3)		ER-13	Internally Generated	Unique sequence number for each company establishment within the state. Includes the establishment registration number.
NSIS	Public Water System	9(7)		MS-2	Assigned by State	A seven digit number uniquely identifying a Public Water System within a State.
NEEDS	Authority/Facility No.	X(9)	F2	NE-1	EPA-1	Unique identifier that consists of authority and facility number.
PENS	Establishment No.	X(9)		PM-4	Region	
PENS	Shipper, Manufacturer, User, Distributor No.	X(6)		PM-7	Sample Collection Report	
PENS	Port/Establishment No.	X(10)	F3	PM-8	Region	
PCS	Facility ID No.	X(9)	F4	PR-3	NPDES Application	The facility ID number.

PPIS	Distributor No.	X(6)		PS-8	Same as the FPA Company Number component of the EPA Registration Number.
RAMS	Plant ID No.	X(4)		RA-1	Unique number identifying a plant.
STORET	Facility ID	X(9)	F4	ST-1	
TSCAS	Plant Site No.			TS-18	Same as Prime Station Number.

DATA CATEGORY Sample Station Site
DATA ELEMENT TYPE Site

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Point Identification No.	X(2)		AE-17		
AEROS/NEDS	Source Classification No.	9(8)		AE-18		
AEROS/SAROAD	Site	X(3)		AE-26		
AEROS/SAROAD	Site Abbreviation	X(25)		AE-34		
CDS	Emission Point No.	9(3)		CD-7	Region	Emission Point Identifier
EDS	Site	X(9)	22	ED-36	SAROAD	
EDS	AAQ Station	X(111)		ED-35	AEROS/SAROAD	State Name, Area Name, County Name, Site from SAROAD.
MSIS	Water Source	X(15)		MS-9	State	Represents the name of a water source used by the public water system.
STORET	Prime Station Number	X(6)		ST-4		
STORET	Secondary Station Number	X(12)		ST-3		

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE Region - EPA

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/SAROAD	Region	X	R2	AE-25		
CDS	Region	9(2)	R1	CD-3	HQ	EPA Region Code.
EDS	Region	X(2)	R1	ED-12	SASD	EPA Region.
ERSS	Region	9(2)	R1	ER-10	System Generated	Region Code.
GICS	Region	9(2)	R1	GI-9	Grant Application	The number of the EPA Region in which the headquarters of the applicant is located (may not be location or facility).
MSIS	Region	X(2)	R1	MS-20	System Generated	The region code of the public water system for which the sample was taken.
PEMS	Region	X(2)	R1	PM-10	Regions	EPA Region Code.
PCS	Region	9(2)	R1	PR-1	Regions	EPA Region Code.
PPIS	Region	X	R2	PS-6		EPA Region Code.
TOXICS	Region	X(2)	R1	TS-17	System Generated	EPA Region Code.
TOXICS	Region	X	R2	TS-15	System Generated	EPA Region Code.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE State

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	State Code	X(2)	S2	AE-7		
AEROS/SAROAD	State Code	X(2)	S2	AE-24		
CDS	State Code	9(2)	S4	CD-4	SAROAD	Two position numeric state code - SAROAD.
CDS	State Abbreviation	X(2)	S7	CD-17	Internally Generated	Two position alphabetic state code used by the U.S. Post Office.
EDS	State Code	X(2)	S2	ED-7	NEDS/SAROAD	AEROS State code.
EDS	State Name	X(20)		ED-21	FPC 67/AEROS	
ERSS	State Name	X(2)		ER-6	Registration Form	Standard two character state name abbreviation of the home office or headquarters where the company is located.
ERSS	State Code	9(2)		ER7	Internally Generated	
GICS	State Abbreviation	A(2)	S8	GI-5	State Project Priority List Organization Application	The FIPS/US Postal Service abbreviation for the name of the state in which the headquarters of the applicant is located (applicant not site).
GICS	Applicant/Project State	A(4)	S8	GI-20	Grant Application	Identifies both the applicant's state and project site state for research and demonstration applications.
MSIS	State Code	9(2)		MS-7	State	Represents a two digit code of the state responsible for the Public Water System.
MSIS	State Code	A(2)	S7	MS-12	State	Represents the postal code of the state used with the Public Water System plant or owner address.

NEEDS	State Code	X(2)	S3	NE-4		Two-digit number will indicate the name of the state in which the facility is physically located (FIPS-4).
PEMS	State Abbreviations	X(2)	S9	PM-16	Sample Collection Rpt.	
PEMS	State Code	X(2)	S1	PM-1	Region	
PCS	State Code	A(2)	S1	PR-8	Region	FIPS Alpha State Code
PPIS	State and City Name	X(45)		PS-4		
Pesticides	State Code	X(2)	S3	PS-7		FIPS State Code
RAMS	State Code	X(2)		RA-8		
STORET	State Abbreviation	X(2)	S6	ST-8		
STORET	State Code	X(2)	S4	ST-13		
TOXICS	State Abbreviation	A(2)		TS-6	Chemical Substances Inventory Report	Name of the State where the chemical substances are manufactured or processed.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE County

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	County/City Code	X(4)	A2	AE-5		City or County Code, depending on flag
AEROS/NEDS	County/City Name	X(28)		AE-3		Name of City, County, or State, depending on flag.
AEROS/SAROAD	County/City Code	X(4)	A2	AE-5		City or County Code, depending on flag
AEROS/SAROAD	County/City Name	X(28)		AE-3		Name of City, County, or State depending on flag.
AEROS/SAROAD	County Name	X(15)		AE-6		
AEROS	County Identifier	X(6)		AE-4		Number of counties making up the ACQR
CDS	County Code	9(4)	A2	CD-5		Four digit numeric county identifier derived from the SAROAD manual.
CDS	County Name	A(17)		CD-16	Internally Generated	County name generated by the edit from the county code.
EDS	County Code	X(6)	A1	ED-6	AEROS	AEROS County Code
EDS	County Name	X(25)		ED-22	FPC 67/AEROS	
GICS	County Name	A(15)		GI-7	State Project Priority List or Grant application	The name of the county (1) in which the project site is located, or (2) which the project serves.
MSIS	District	X(2)		MS-39	State	A code identifying an administrative district within a region (can be county)
MSIS	Region-State	X(2)		MS-38	State	A code identifying an administrative region of the state.
MSIS	County Code	9(3)	A3	MS-36	State	Represents the county code of the county in which the

						service area of a Public Water System is located (FIPS-6).
NEEDS	County Code	X(2)	A3	NE-4	EPA 1	The code of the county in which the facility is physically located (FIPS-6).
NEEDS	County Name	X(20)		NE-9	EPA 1	The name of the county in which the facility is located.
PCS	County Name	X(20)		PR-6	City Master File	
PCS	County Code	X(3)	A3	PR-7	Region	Defines the county in which a facility is located (FIPS).
RANS	County Code	X(4)		RA-9		Four position number allocated to that county within the state.
STORET	County Code	X(3)	A3	ST-12		
TOXICS	County Name	A(28)		TS-5	Chemical Substance Inventory Report	Name of the county where the chemical substances are manufactured or processed.
TOXICS	County Code	X(3)		TS-16	Internally Generated	

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE City

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	County/City Code	X(4)	A2	AE-5		City or County Code depending on flag
AEROS/NEDS	County/City Name	X(28)		AE-3		Name of City, County or State depending on flag. City must have population of over 2,400.
AEROS/SAROAD	County/City Code	X(4)	A2	AE-5		City or County Code depending on flag.
AEROS/SAROAD	County/City Name	X(28)		AE-3		Name of City, County or State depending on flag. City must have population of over 2,400.
CDS	City Code	9(4)	A2	CD-9	Region	Four digit numeric city identifier.
CDS	City Name	X(15)		CD-12	Region	City location of the facility.
EDS	City Code	X(6)	A1	ED-5	NEDS/SAROAD	AEROS City Code
EDS	Town and Zip Code	X(30)		ED-24	FPC 67	
ERSS	City Name	X(20)		ER-5	Registration Division	City name where the home office or headquarters of the company is located.
ERSS	City Name	X(20)		ER-15	Establishment Registration Forms	City of the establishment
GICS List or Grant application	City Name the headquarters of the applicant is located.	A(17)		GI-6	State Project Priority	The name of the city in which
MSIS	City Name	X(15)		MS-3	State	Represents the name of the city or community serviced by the Public Water System.
MSIS	City Name	X(30)		MS-6	State	Represents the owner or plant city name
NEEDS	City Name	X(25)		NE-8	EPA 1	The name of the city or town in which the facility is physically located.

PERS	City Name	X(20)		PM-2	Region	Establishment City.
Permits	City Code	X(5)	A4	PR-4	Region	City Master File
PCS	City Name	X(20)		PR-5	Region	The name of the city where the facility is located.
PPIS	City Name and State	X(45)		PS-4		
RANS	City Code	X(4)		RA-10		City in which plant is located.
STORET	City Code	X(5)		ST-11		(User Option)
TOXICS	City Name	A(20)		TS-4	Chemical Substances Inventory Report	Name of the city where the chemical substances are manufactured or processed.
TOXICS	City Code	X(5)		TS-15	Internally Generated	

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE Address

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Name and Address	X(40)		AE-1		Establishment name and address
AEROS/SAROAD	Site Abbreviation	X(25)		AE-34		
AEROS/SAROAD	Site Address	X(41)		AE-35		
CDS	Street Address	X(20)		CD-11	Region	Actual location of the facility
EDS	Post Office	X(25)		ED-20	FPC 67/AEROS	
EDS	Street and Number	X(30)		ED-19	FPC 67	
ERSS	Street Name	X(40)		ER-4	Registration Division	Street address of the home office or headquarters of the company submitting the application.
ERSS	Street Name	X(40)		ER-14	Establishment Registration Form	Street address of establishment.
GICS	Street Address	X(30)		GI-13	State Project Priority List or Application Form	Mailing address of the applicant.
MSIS	Street Address	X(30)		MS-5	State	The owner or plant street address
PEMS	Smud Address	X(23)		PM-5	Region	Shipper, Manufacturer, User distribution address
PPIS	Street Address	X(50)		PS-3		
RAMS	Street Address	X(28)		RA-11		The number and street of the plant.
TOXICS	Plant Address	X(50)		TS-3	Chemical Substances Inventory Report	Address of the plant where the chemical substances are manufactured.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE Zip Code

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
CDS	Zip Code	9(5)		CD-13	Region	Five digit numeric postal zone.
EDS	Zip Code	X(5)		ED-23	FPC 67	
ERSS	Zip Code	9(5)		ER-9	Internally Generated	Standard mailing zip code of the headquarters or home office of the company.
GICS	Zip Code	9(5)		GI-14	Grant Application	U.S. Postal Service zip code.
MSIS	Zip Code	9(5)		MS-16	State	Represents the zip code of the Public Water System plant or owner address.
NEEDS	Zip Code	X(5)		NE-10	EPA 1	The official U.S. Post Office zip code for the facility location.
PPIS	Zip Code	9(5)		PS-5		
TOXICS	Zip Code	9(5)		TS-7	Chemical Substances Inventory Report	Zip Code of plant where the chemical substances are manufactured or processed.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE SMSA

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS	SMSA Code	X(4)		AE-10		Standard Metropolitan Statistical Area Code.
AEROS	SMSA Name	X(41)		AE-11		Standard Metropolitan Statistical Area Name.
EDS	SMSA Code	X(4)		ED-8	SAROAD	Standard Metropolitan Statistical Area Code.
EDS	SMSA Name	X(41)		ED-9	AEROS	Standard Metropolitan Statistical Area Name.
GICS	SMSA				Zip Code	GICS provides SMSA Data on output reports by using a Zip Code link to an external table that contains the Zip Code and SMSA.
NEEDS	SMSA Code	X(5)		NE-5	EPA 1	Standard Metropolitan Statistical Area Code. Place "No" if area is not in SMSA.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE UTM

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	UTM			AE 19,20,21		UTM Zone, vertical and horizontal coordinates.
AEROS/SAROAD	UTM			AE 30,31		UTM Zone, northing and easting coordinates.
EDS	UTM			ED 14,15,16		UTM Zone, northing and easting coordinates.
RANS	UTM			RA 6,7,1		Grid Number, vertical and horizontal coordinates.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE Longitude and Latitude

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/SAROAD	Longitude and Latitude	X(8),X(7)		AE 28,29		Direction, degrees, minutes and seconds.
EDS	Longitude and Latitude	X(8),X(7)		ED 17,18	Internally Generated	Degrees, minutes, seconds.
MSIG	Longitude and Latitude	X(7),X(7)		MS 10,11	State	Degrees, minutes, seconds.
STORET	Longitude and Latitude	X(8),X(7)		ST2		Degrees, minutes, seconds.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE AQCR, AQMA

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/SAROAD	AQCR-Number	X(3)		AE-9		Air Quality Control Region No.
AEROS/SAROAD	AQCR - Name	X(55)		AE-8		Control Region Name
CDS	AQCR-Number	X(3)		CD-8	SAROAD/NEDS	Air Quality Control Region No.
EDS	AQCR-Number	X(3)		ED-13	NEDS	Air Quality Control Region No.
EDS	AQMA-Name	X(40)		ED-11		Air Quality Maintenance Area Name (Not Presently in Use)
EDS	AQMA-Number	X(6)		ED-10		AQMA Code (Not Presently in Use).

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE River Basins

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
GICS	Facility River Basin Code	9(6)	B1	GI-10	Grant Application	A numeric code approved by EPA, designating the hydrologic planning area in which the discharge point of a wastewater treatment facility is located.
GICS	303(E) Basin Code	X(5)	B2	GI-17	Office of Water Planning	A numeric and alpha code, approved by EPA, designating the 303(E) planning area in which the source of a discharge is physically located. The areas are determined by each state and are used in preparing 303(E) Basin Plans.
MSIS	Plant Basin Number	9(4)	B3	MS-24	State	Represents a code indicating a river or lake from which water is obtained by the Public Water System.
NEEDS	Basin Code	X(4)	B5	NE-6	EPA 1	Basin Codes developed by EPA.
PCS	River Basin Code	X(4) or X(6)	B5 or B1	PR-12	Region	Used to identify the river basin receiving discharge.
STORET	River Basin Code	X(6)	B1	ST-10		
STORET	Hydrologic Index	X(15) to X(112)	B4	ST-5		A numerical code which identifies the location of a sampling station on a defined river system by defining its distance and relation to the mouth of a river system.

DATA CATEGORY Facility Geographical Location
DATA ELEMENT TYPE Other

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/SAROAD	Area	X(4)		AE-23		
AEROS/SAROAD	Site	X(16)	Z1	AE-33		
EDS	Site	X(9)	Z2	ED-36	SAROAD	
EDS	Surface Data	X(15)		ED-33	Waldel/Geomet	Location of Meteorological Surface Data
EDS	Mixing Height	X(15)		ED-34	Waldel/Geomet	Location of Meteorological Mixing Height Data.
FMS	Location Code	9(9)	Z4	FM-1		Location related to a grantee or contractor.
GICS	Congressional District	X(10)	Z6	GI-8	Grant Application	The number of the congressional district in which the applicant is located.
NEEDS	Facility Location	X(8)	Z3	NE-4	EPA 1	
NEEDS	Congressional District	X(9)		NE-7	EPA 1	The number of the major congressional district served.
PEMS	Establishment Site No.	X(10)		PM-17	Sample Collection Report	
PCS	Location Code	X(7)	Z5			
RAMS	Zone	9(2)		RA-5		Stack Location
RAMS	Area	X(4)		RA-5		Unique number identifying area in which point resides.

DATA CATEGORY Parameter Unit Identifiers
DATA ELEMENT TYPE Unit

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Heat Units	X		AE-22		
AEROS/NEDS	Units	9		AE-22		
AEROS/NEDS	Unit System	9	U1	AE-36		Indicates if english or metric unit system is selected.
AEROS/SAROAD	Units	X(2)		AE-27		Code for standard units.
EDS	Particulate Matter Units	X(12)		ED-28	SASD	Unit of measure of primary standard condition for particulate matter.
EDS	SO2 Units	X(12)		ED-27	SASD	Units of measure of primary standard condition for SO2.
EDS	Pollutant Units	X(12)		ED-32		Units of measure of amount of pollutant measured - reserved for future inclusions.
EDS	Fuel Units	9(2)	U3	ED-31	FPC-67	Units of measure of quantity of fuel to be used - reserved for future inclusion.
EDS	Units of measure	X(20)	U3	ED-30	SASD	Name of the code for units of measure
EDS	Code	X(2)	U3	ED-29		Code for units of measure.
EDS	Units of measure	(Repeating)		ED-37	SASD	Identifies unit of measure for each fuel type.
ERSS	Units of measure	A	U4	ER-20	Pesticides Report	Indicates the unit of measure of the product.
MSIS	Chemical Units	X(6)		MS-1	State	Name of the units used to measure contaminant level in drinking water.

MSIS	Unit System	A	U5	MS-8	State	Represents a flag indicating whether the english (gallons/fahrenheit) or metric (liters/celsius) system of measuring is used to record plant capacities and air temperature by the Public Water System.
PCS	Discharge Monitoring Report Unit	X	U2	PR-9	Discharge Monitoring Report	Designates whether the reporting period is based on months or days.
STORET	Parameter Code Units	9(5)		ST-6		Analysis performed including the units of the item reported from the laboratory.

DATA CATEGORY Quality Assurance Codes
DATA ELEMENT TYPE

SYSTEM	NAME	PICTURE	CODE	SHEET	SOURCE	DEFINITION
AEROS/NEDS	Regulatory	X(2)	G1	AE-12		The year the regulatory information was last updated.
AEROS/NEDS	Compliance	X(6)	G3	AE-13		Date compliance information was updated.
AEROS/NEDS	Emissions	X(2)	G1	AE-14		The year the emissions information was last updated.
AEROS/NEDS	Records	X(2)	G1	AE-15		The year when data was last updated.
CDS	Date last updated	9(6)	G3	CD-2		The most current date on which an input transaction was added or changed for any particular facility or the master file.
GICS	Activity Date	9(6)		COM	Internally Generated	This is the date a GICS record was last updated.
MSIS	Records	9(5)	G2	MS-22	Internally Generated	This is the date an MSIS record was updated. It is used on the MSIS history file to identify changed data elements according to the date of the change. The date is in Julian Format.
PEMS	Date Collected	X(6)	G3	PM-19	Sample Collection Report	
STORET	Quality Assurance	X(1)	G4	COM		
TOXICS	Entry Date	9(5)	G2	TS-9	CAS	Date of last entry in Julian Format.
TOXICS	Last Activity Date	9(5)	G2	TS-10	CAS	Date of last activity in Julian Format.

APPENDIX B
CODING SCHEMES

APPENDIX B

TABLE OF CONTENTS

INTRODUCTION

The coding schemes presented in this appendix were identified during the review of the 15 representative systems as part of the quantitative analysis of the need for a data management and standardization program. Three types of information is presented for each identified coding scheme. They are:

- Code No. - The Code Number is a project team assigned number used to identify the specific coding scheme. The numbers have an alphabetic prefix to identify the coding scheme category and a sequentially assign number. The code number is used in the Data Dictionary in Appendix A to reference the specific coding scheme used for that data element.

- Field - Information is provided in this column when specific information about the data element is provided by specific character locations in the coding scheme. For example, the first two characters in a location code may identify the state and the last three characters, the county.

- Type - This column contains the name and/or description of the specific coding scheme.

CITY AND COUNTY CODES

Code No.	Field	Type
A1	1-2	State Code - SAROAD
	3-6	County or County Equivalent AEROS Code or AEROS City Code
A2		The four digit County or City code explained above.
A3		A three digit sequential County code by State (FIPS PUB 6-1).
A4		EPA City Master File

RIVER BASIN CODES

Code No.	Field	Type
B1	1-2	Major River Basin
	3-4	Minor River Basin
	5-6	River Sub-Basin
B2	1-2	303(E) Basin Code - A numeric code identifying a 303(E) basin within a State. This code is used in conjunction with data element "Applicant State Abbreviation," in order to identify the State.
	3-4	Discharge River Basin Segment - A numeric code identifying the specific segment within a 303(E) basin.
	5	Segment Class - An alpha code identifying the limitation class applicable to the specific 303(E) basin segment; that is, whether it is Effluent limited ("E") or Water Quality Limited ("W"). This data element should be entered for all WWT facilities including those which utilize land application techniques for treated effluent. Legitimate entries are: Positions 1-2 - "00" - "99" Positions 3-4 - "00" - "99" Position 5 - "E", "W" or blank if unknown
B3	1-2	Geographic Region Number
	3-4	Specific Basin Number - Sequential
B4	1-2	Major Basin Code
	3-7	Minor Basin Code
	5-7	Terminal Stream No.
	Var.	Indices which define the direction and level of stream flow
	Var.	Mileages which define the distances between two confluences in the river system
	Var.	A code which identifies the stream level on which the point is located

B5	1-2	Major Basin Code
	3-4	Minor Basin Code

COMPANY CODES

Code No.	Field	Type
C1	1-9	IRS Tax Code
	10-11	Assigned Locally

FACILITY CODES

Code No	Field	Type
F1	1-2	Numeric FIPS State Code
	3-7	Zip Code
	8-9	Authority
	10-11	Facility
F2	1-2	The FIPS-5 Codes for States and Territories of the United States
	3-6	An authority within the State or Territory
	7-9	An individual facility within the Authority or "No-Number"
F3	1-2	The two digit alphabetic code for States and Territories
	3-6	A sequential number identifying a port of entry.
F4	1-2	2 character FIPS alpha State code
	3-8	6 digit sequence number
	9	1 character check digit
F5		May use NEDS or sequential
F6		FPC Plant Code

DATE CODES

Code No.	Type
G1	Year
G2	Date - Julian
G3	Date - Gregorian
G4	A - I - Very Bad to Good

EPA REGION CODES

R1	R2			
01	1	EPA Region I	-	Boston
02	2	EPA Region II	-	New York
03	3	EPA Region III	-	Philadelphia
04	4	EPA Region IV	-	Atlanta
05	5	EPA Region V	-	Chicago
06	6	EPA Region VI	-	Dallas
07	7	EPA Region VII	-	Kansas City
08	8	EPA Region VIII	-	Denver
09	9	EPA Region IX	-	San Francisco
10	0	EPA Region X	-	Seattle

CODES FOR STATES AND TERRITORIES OF THE UNITED STATES

	S1	S2	S3
Alabama	AL	01	01
Alaska	AK	02	02
Arizona	AZ	03	04
Arkansas	AR	04	05
California	CA	05	06
Colorado	CO	06	08
Connecticut	CT	07	09
Delaware	DE	08	10
District of Columbia	DC	09	11
Florida	FL	10	12
Georgia	GA	11	13
Hawaii	HI	12	15
Idaho	ID	13	16
Illinois	IL	14	17
Indiana	IN	15	18
Iowa	IA	16	19
Kansas	KS	17	20
Kentucky	KY	18	21
Louisiana	LA	19	22
Maine	ME	20	23
Maryland	MD	21	24
Massachusetts	MA	22	25
Michigan	MI	23	26
Minnesota	MN	24	27
Mississippi	MS	25	28
Missouri	MO	26	29
Montana	MT	27	30
Nebraska	NE	28	31
Nevada	NV	29	32
New Hampshire	NH	30	33
New Jersey	NJ	31	34
New Mexico	NM	32	35
New York	NY	33	36
North Carolina	NC	34	37
North Dakota	ND	35	38
Ohio	OH	36	39
Oklahoma	OK	37	40
Oregon	OR	38	41
Pennsylvania	PA	39	42
Rhode Island	RI	41	44
South Carolina	SC	42	45
South Dakota	SD	43	46
Tennessee	TN	44	47
Texas	TX	45	48
Utah	UT	46	49
Vermont	VT	47	50
Virginia	VA	48	51
Washington	WA	49	53
West Virginia	WV	50	54
Wisconsin	WI	51	55
Wyoming	WY	52	56
Puerto Rico		40	43
American Samoa		53	
Guam		54	
Virgin Islands		55	

S4 Codes used in "S2" plus:
Trust Territories 56

S5 Codes used in "S1" plus the numeric codes shown below:

S6 Codes used in "S3" plus the alphabetic codes shown below:

Name	State Code (Numeric)	Country Code (Alpha)	Post Office Abbr.
American Samoa	60	AQ	
Canal Zone	61	PQ	CZ
Canton and Enderbury Islands	62	EQ	
Guam	66	GQ	GU
Johnston Atoll	67	JQ	
Midway Islands	71	MQ	
Puerto Rico	72	RQ	PR
Ryukyu Islands, Southern	73	YQ	
Swan Islands	74	SQ	
Trust Territories of the Pacific Islands	75	TQ	
U.S. Miscellaneous Carribbean Islands	76	BQ	
U.S. Miscellaneous Pacific Islands	77	IQ	
Virgin Islands	78	VQ	VI
Wake Island	79	WQ	

S7 Codes used in "S1" plus:

American Samoa	AS
Guam	GU
Virgin Islands	VI
Trust Territories	TT
Puerto Rico	PR

S8 Codes used in "S1" with the following changes:

Trust Territories of the Pacific Islands	PI
Puerto Rico	PR
American Samoa	SA
Virgin Islands	VI
Guam	GU
Foreign Country	FC
Northern Mariano	CQ

S9 Codes used in "S1" plus:

Puerto Rico	PR
Guam	GU
Virgin Islands	VI
Africa	AF
Asia	AS
Australia	AU
Europe	EU
North America	NA
South America	SA

UNIT CODES

U1 1 = metric 2 = english

U2 M = months D = days

U3 00 -- No emission limit
01 -- %S - all fuels
02 -- %S - each fuel
03 -- #SO_2/MMBtu - all fuels
04 -- #SO_2/MMBtu - each fuel
05 -- #S/MMBtu - all fuels
06 -- #S/MMBtu - each fuel
07 -- ppm SO_2 - emission
08 -- ppm SO_2 - ambient air
09 -- #SO_2/hr

11 -- #Particulate/MMBtu
12 -- Grains/SCF
13 -- Grains/SCFD
14 -- #Particulate/hr
15 -- #Particulate/1000#Stack Gas
16 -- #Particulate/1000#Bagasse
17 -- Ambient Air Standard
18 -- % Control

21 -- #NO_x/MMBtu - all fuels
22 -- #NO_x/MMBtu - each fuel
23 -- ppm NO_x - emission
24 -- ppm NO_x - ambient air
25 -- #NO_x/hr

U4 L = pounds G = gallons

U5 E = english M = metric

Code Number	Units
01	micrograms/cubic meter (25°C, 1013 millibars)
02	micrograms/cubic meter (0°C, 1013 millibars)
03	nanograms/cubic meter (25°C, 1013 millibars)
04	nanograms/cubic meter (0°C, 1013 millibars)
05	milligrams/cubic meter (25°C, 1013 millibars)
06	milligrams/cubic meter (0°C, 1013 millibars)
07	parts per million (volume/volume)
08	parts per billion (volume/volume)
09	COHS/1,000 linear feet
10	RUDS/10,000 linear feet
11	meters/second
12	miles/hour
13	knots
14	degrees - compass (1° - 360°)
15	degrees - Fahrenheit
16	millibars
17	degrees - Centigrade
18	langleys (gram-calories/square centimeter)
19	percent relative humidity
20	microns
21	inches (rainfall)
22	inches (mercury)
23	millicalories/square centimeter/minute
24	miles (visibility)
25	langleys/minute
26	degrees - Rankine
27	Beta scanner
28	degrees Centigrade/100 meters
29	millimeters (rainfall)
30	picocuries/cubic meter
31	microcuries/cubic meter
32	picocuries/square meter
33	microcuries/square meter
34	picocuries/cubic centimeter
35	picocuries/gram
36	calories/square centimeter/hour
37	degrees - Kelvin
40	parts per hundred million
41	milligrams SO_3/100 square centimeters/month[a]
42	RUDS/1,000 linear feet
43	grams/square meter/month[a]
44	micrograms/square mile/month[a]
45	tons/square mile/day
46	grams/square meter/day
47	micrograms/square meter/day
50	Number of threshold levels
51	% loss in reflectance/month
52	microns/week
53	number defects/7.7 in^2month
54	particules/Mm^2/week

55	NBS color difference units
56	microns/year
61	pH scale
62	milligrams/liter
67	micro equivalence
69	micro siemens/centimeter
70	milligrams F/100 square centimeters/day
71	micrograms F/100 square centimeters/day
80	milligrams SO_3/100 square centimeters/day
81	micrograms SO_2/square centimeter/day
82	micrograms SO_2/square meter/day
90	tons/square mile/month[a]
91	milligrams/square centimeter/month[a]
92	micrograms/cubic meter/month[a]
93	grams/square meter/month[a]
94	pounds/square mile/month[a]
97	micrograms/square centimeters/30 days
98	milligrams SO_4/square centimeters/30 days
99	milligrams/square centimeters/30 days

[a]On a calendar month basis.

LOCATION CODES

Z1 Code consists of:

Field Length	
3	AQCR Code
2	State Code
4	Area Code
3	Site Code
1	Agency Code
2	Project Code
1	Blank

Z2 Code consists of:

Field Length	
2	AEROS State Code
4	Area Code
3	Site Code

(same codes as used above)

Z3 Code consists of:

Field Length	
2	State Code FIPS-5
2	County Code FIPS-6
4	Place Code

Z4 Nine digit code which identifies state, county and city

Z5 Alpha state code (FIPS) and the city code (EPA City Master File)

Z6 Special Codes are:

00	-	No Congressional District
66	-	Los Angeles
77	-	Chicago
88	-	New York City
99	-	All Congressional District in the State