EPA/620/R-94/017
June 1994

EMAP

INFORMATION MANAGEMENT STRATEGIC PLAN: 1993 - 1997

Environmental Monitoring and Assessment Program
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC 20460

Abstract

The *Information Management Strategic Plan* for the Environmental Monitoring and Assessment Program (EMAP) describes how EMAP information will be managed from field sampling through the delivery of products to the user. This Plan cuts across the entire EMAP program and explains the process for developing an information management infrastructure which will enable EMAP to achieve its long-term objectives. This Plan provides a strategy for meeting hardware, software, documentation support, and system designs that will meet EMAP Resource and Coordinating Group requirements. Users of EMAP data and information are discussed in the Plan as well as the process and technology with which they can access EMAP data and other data sources.

Key words:

environmental monitoring, environmental assessment, information management, information systems, systems architecture, systems engineering, database management system, USEPA-EMAP, information resource management, geographic information system

Preferred citation:

Shepanek, R., 1994. EMAP Information Management Strategic Plan: 1993-1997. EPA/620/R-94/017, Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program.

Notice:

This report was submitted in fulfillment of contract # DW89935335 to Battelle Pacific Northwest Laboratories under the sponsorship of the United States Environmental Protection Agency. This report covers a period FY 1993 to FY 1997 and work was completed as of March, 1994.

Disclaimer

Preface

The Environmental Monitoring and Assessment Program (EMAP) is a nationwide program to monitor and assess the state of ecological resources in the United States. It will collect, organize, and provide for the analysis, assessment and reporting of large volumes of spatial and temporal ecological data and information. This Plan, *the EMAP Information Management Strategic Plan: 1993-1997*, provides the vision, scope, approach, and resource requirements for EPA management to guide the information management aspect of the program.

This Plan has been developed through a systems engineering process similar to that proposed for the development of EMAP IM operating systems. The first step involved a facilitated design session involving management, users, and developers. This produced a rough version of the Plan for review. Three iterative versions then underwent broader review, with each subsequent version having an increasingly wider audience. A Peer Review conducted in March 1993 provided detailed comment and guidance from experts in computer systems and the environmental sciences outside of EPA.

Since its inception, this Plan has changed both in structure and content. Nonetheless, it has been a common departure point for receiving constructive feedback and for helping to form consensus on "mission and direction" for EMAP Information Management. It is intended to be a living document which will serve as an introduction tool for new staff, and will be a baseline for incorporating changing requirements and technology in EMAP IM systems through subsequent updates.

Principal Investigator: Robert Shepanek, Ph.D.
EMAP Information Management Coordinator
U.S. EPA (8205)
401 M Street, S.W.
Washington, DC 20460
Telephone: (202) 260-3255
Internet: shepanek.robert@epamail.epa.gov

Acknowledgements

The Environmental Monitoring and Assessment Program (EMAP) Information Management Strategic Plan went through two phases: the pre-peer review phase to March 1993, and the peer review/post-peer review phase which resulted in this final document. While many EPA staff, peer reviewers, and consultants provided valuable insight and technical expertise to each phase, special acknowledgement is accorded the following individuals who were principally involved throughout the evolution of this Plan.

Within EMAP, key participants included: Don Worley, Linda Kirkland, Denice Shaw, Rick Linthurst, Dick Latimer, Mason Hewitt, Chuck Liff, Mark Tooley, Sidney Draggan, Dan McKenzie, Dwight Clay, and Gary Collins.

The members of the Peer Review team who were instrumental in shaping the final document included: Chairperson, Maria Zemankova, National Science Foundation; Eliot Christian, U.S. Geological Survey; Lois Delcambre, Oregon Graduate Institute of Science and Technology; Michael Folk, National Center for Supercomputing Application; Sara Graves, University of Alabama in Huntsville; Bruce Gritton, Monterey Bay Aquarium Research Institute; Richard Haeuber, Sustainable Biosphere Initiative; Eric Rodenburg, World Resources Institute; John Rumble, National Institute of Standards and Technology; and Arie Shoshoni, Lawrence Berkeley Laboratory.

Consultants included: Jim Thomas, Robert Isaak, Bruce Kissinger, Kathy Moore, and Jim Brown of Battelle Pacific Northwest Laboratories; Jeff Frithsen of VERSAR; Jeff Rosen of American Management Systems; Jon Clark and Jane Lovelace of Technology Planning & Management Corp.; Bob Booher of Dyncorp Viar; and Rod Slagle of Lockheed.

TABLE OF CONTENTS

EMAP Information Management Strategic Plan: 1993 - 1997

APPENDIX

EMAP

INFORMATION MANAGEMENT STRATEGIC PLAN: 1993 - 1997

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

EMAP Commitment

The Environmental Monitoring and Assessment Program (EMAP) represents a long-term commitment by the U.S. Environmental Protection Agency (EPA) to assess and periodically document the status and condition of the Nation's ecological resources. EMAP's goal is to monitor and assess the condition of those resources and to contribute to decisions on environmental protection and management.[1] EMAP will monitor indicators of the condition of our Nation's ecological resources to respond to the growing demand for information characterizing the condition of our environment and the type and location of environmental changes. The EMAP project, when implemented, will provide the ability to integrate information from multiple sources across environmental media within the context of the National Information Infrastructure[2] and provide assessments of environmental impacts across ecological resource categories.

Strategic Plan

This EMAP Information Management (IM) Strategic Plan outlines the implementation approach for development of workable information systems that will meet the requirements of EMAP scientists, information managers, and government decision makers. This Plan provides EPA management with the scope, the foundation of requirements and guiding principles, the management and technical approach, and resource requirements for EMAP information systems development. This Plan also provides EPA environmental scientists and collaborators with a common set of expectations, and the EPA information management staff with a systems engineering road map that can be used to direct activities over the next 5 years.

EMAP IM Vision

By 1997, EMAP is expected to monitor and assess the current status and geographic distribution of such ecological resources as estuaries, lakes, streams, wetlands, forests, grasslands, and deserts. By integrating information from several government agencies, assessments and reports will become available on the proportions of these resources that are degrading or improving, their location and extent, and their rate of change. Reported ecological conditions and geographical coverage shall be available with known confidence.

EMAP Program Objectives	EMAP IM Objectives
Estimate the current status, trends, and changes in selected indicators of the condition of the Nation's ecological resources on a regional basis with known confidence.	• Assist in structuring, developing, maintaining, operating, and/or deploying: - Resource Group databases for research, monitoring, and analysis; - Access mechanisms to the databases; - Manipulation mechanisms (algorithms); and - Display mechanisms required for research, monitoring and analysis.
Estimate the geographic coverage and extent of the Nation's ecological resources with known confidence.	• Assist in structuring, developing maintaining, operating, or deploying: - Access mechanism to EMAP Geographic Reference Data (GRD); - Manipulation mechanisms (algorithms); and - Display mechanisms required to analyze this information.
Seek associations between indicators of natural and anthropogenic stresses and indicators of condition of ecological resources.	• Ensure a distributed database structure that allows integration of information internally, and with external sources allowing responsibility for the data to reside with the "owners." • Assist in the development, maintenance, operation, and deployment of cross-cutting tools.
Provide annual statistical summaries and periodic assessments of the Nation's ecological resources.	• Ensure that all data, metadata, and information required to meet the objective are available, in some cases by leveraging the central node of the database system. • Assist in the development, maintenance, operation, and deployment of access mechanisms and display mechanisms useful in meeting this objective.

Figure ES-1. EMAP IM objectives.

By 2002, EMAP plans to expand its monitoring and assessment capability to include information from virtually all applicable Federal sources. Degrading and improving trends will be identified through specific indicators of how adversely affected ecosystems are responding to control and mitigation programs.

By 2012, EMAP should be a focused, mature program that is integrated with scientific information from key national and international sources. Trending information of all ecosystems will be available for use in making decision concerning national and international environmental management and economic development.

EMAP Information Management (IM)

EMAP IM is the vehicle with which the total EMAP program manages data and information. EMAP Program objectives and corresponding EMAP IM objectives are shown in Figure ES-1. Specifically, EMAP IM will provide effective and extensible information management systems for assessing and reporting on the condition of the Nation's ecological resources. IM will provide the tools for turning raw scientific measurements (data) into useful information products that will serve to guide actions to improve the quality of our environment.

EMAP IM will be an environmental "information highway" allowing analysis of EMAP data across heterogeneous networks of personal and scientific computers by 1997. EMAP IM systems will provide interfaces for new analysis methods and integration of related scientific information across a "national information infrastructure." Scientists will have access to data distributed across a national network[2] and the computational tools needed to facilitate analyses. EMAP IM systems will provide access to data, metadata, and processed information for a diverse set of users through a network of dispersed databases.

The EMAP Enterprise

The technical and management approaches discussed in this Plan require an explanation of the term "enterprise". The EMAP Enterprise, when fully developed, will include Resource Groups, Coordinating Groups, EPA and non-EPA related programs, and administrative activities that support the program. Collectively, EMAP as an enterprise is made up of these user groups, each of which have differing but complimentary missions to perform in order for EMAP to accomplish its objectives. EMAP IM will eventually involve all these users and will consider their differing requirements for information management systems.

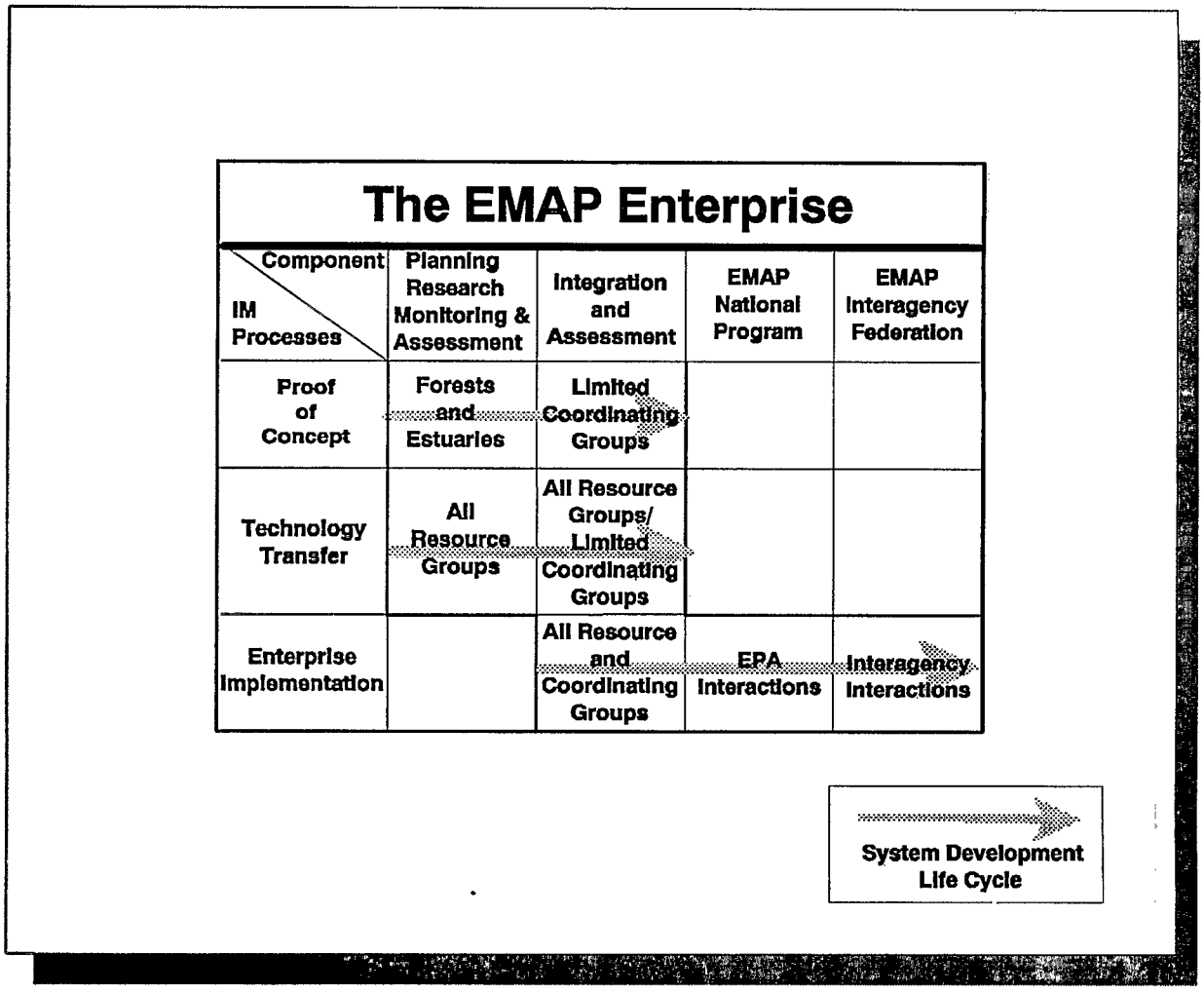

Figure ES-2. The EMAP enterprise.

EMAP is and will continue to be an evolving enterprise with diverse user requirements. Currently, the most "mature" group of users are the Resource Groups. For several years the Resource Groups have been actively engaged in indicator research, collecting data, analyzing data, and publishing results within their ecological areas of responsibility. Conversely, the aspect of the EMAP enterprise least developed is the relationship of EMAP to other EPA and non-EPA programs. For example, such questions as "What information does EMAP have that would assist NASA in accomplishing the objectives of the Earth Observing System Program?" are difficult to answer at this stage of EMAP development.

Enterprise Components

The size and diversity of the collective EMAP user group, and the evolving nature of EMAP requires a systems engineering approach which is stratified by groups comprised of similar users, with similar missions, and in similar stages of development. These user groups are aggregated into four components by level of maturity. Figure ES-2 shows these user group components with the phased IM processes that are applicable to each component. The cells within this map depict the level of user involvement. The "Enterprise Components" of EMAP are:

- ***Planning, Research, Monitoring, and Assessment*** which includes indicator research, collection and analysis of data, and distribution of data and information related to an ecological area of responsibility. This component is comprised of Resource Group users.
- ***Integration and Assessment*** which provides direction on the composition of EMAP information products that are derived from data available within individual Resource Groups, as well as information products derived from multiple Resource Groups. This component includes Coordinating Groups supporting assessment activity by EMAP Central.
- ***EMAP National Program*** component within EPA which defines the relationship of EMAP to other relevant EPA programs. These programs include the Great Lakes National Program, the Gulf of Mexico Program, the STORET Modernization within the Office of Water, the Gateway/Enviro-facts program within the Office of Information Resource Management and others. This component also includes connections to appropriate EPA administrative programs.
- ***EMAP as partner in an Interagency Federation*** defines the relationship of EMAP to relevant programs outside of EPA

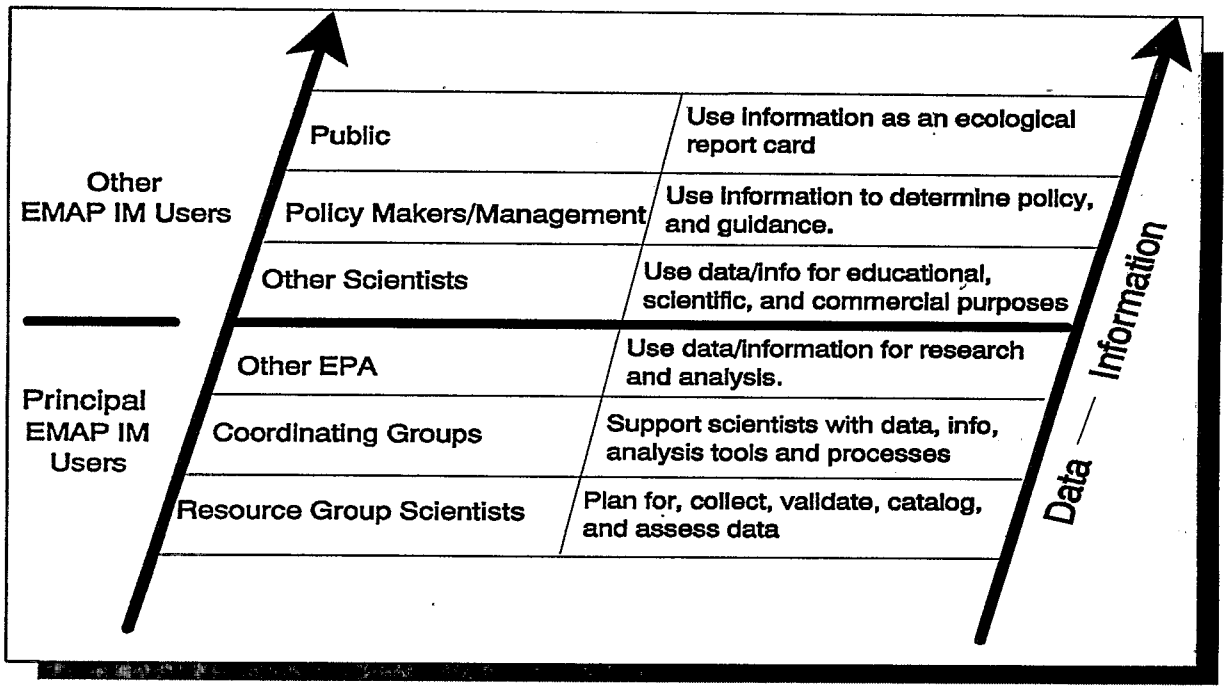

Figure ES-3. EMAP IM users.

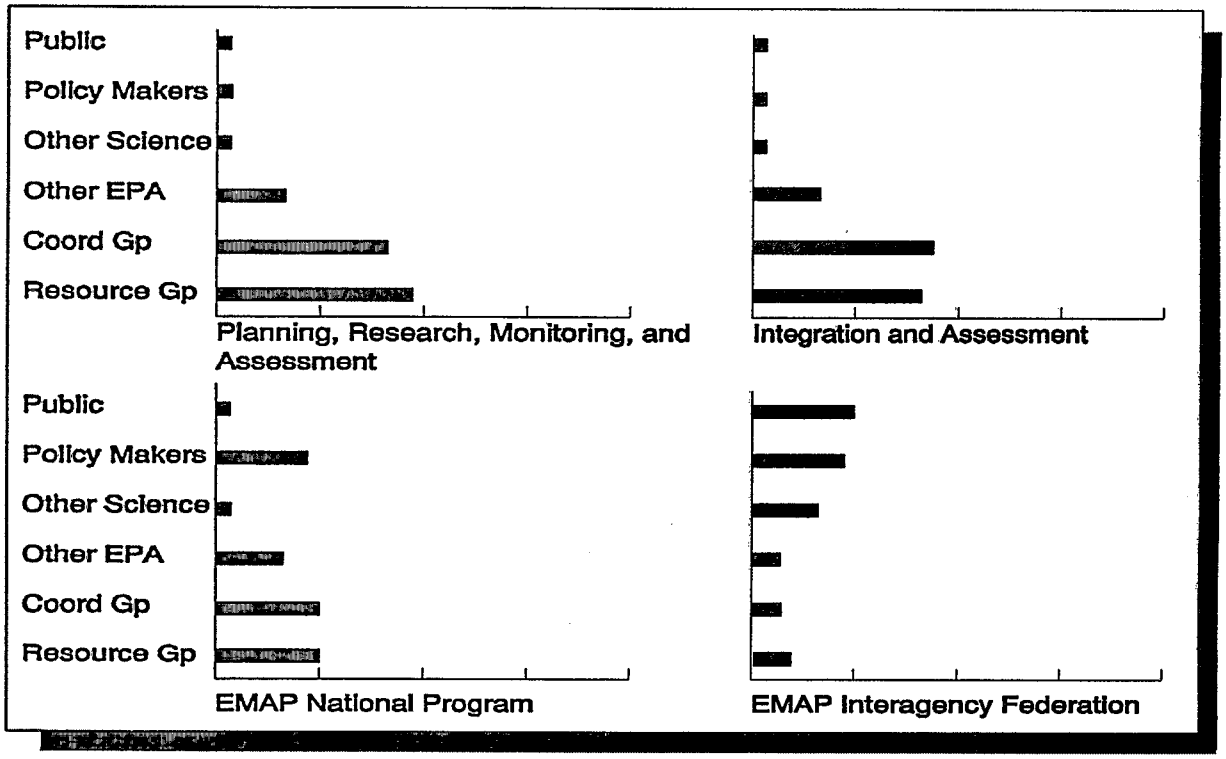

Figure ES-4. Relative user involvement by EMAP IM component.

that are not already included in other EMAP enterprise components by virtue of their cooperative partner status. Examples of the type of programs targeted are NASA's EOS/DIS and the Global Change Research Program.

Each EMAP IM process includes an execution of the systems development life cycle and results with an operational information management system for the designated user. For example, Forests and Estuaries, as well as EMAP Central, will have operational systems upon completion of the Proof-of-Concept (POC). An operational system is the computer hardware, software, databases, and communications links that are required to share data, metadata, and information among users.

EMAP Users

The collective EMAP user community is reflected in Figure ES-3. The first category—and highest priority EMAP user from a strategic and implementation perspective—is the scientific community actively participating in EMAP. These principal users will be scientists in the EMAP Resource and Coordinating Groups, other EPA scientists, and such EMAP Cooperative Partners as the National Oceanographic and Atmospheric Administration (NOAA) and the U.S. Forest Service (USFS). These users have primary responsibility for planning, designing, collecting, managing, researching, validating, analyzing, and interpreting environmental data for a multitude of purposes.

The second user category includes such other EMAP participants as public and private sector scientists and researchers working on behalf of government, academic, and commercial interests; policy makers and government managers who use EMAP information to determine policy and guidance; and the general public and media, all of whom may use EMAP information as an ecological report card.

Eventually, users of EMAP data and information will range from scientists who will extract crosscutting data sets for detailed quantitative analyses to members of Congress, the general public, and the news media who require qualitative information products. A histogram at Figure ES-4 depicts relative differences in user involvement by EMAP Enterprise Components. The make-up of the groups representing the requirements of an EMAP Enterprise component has been structured to exert a leveling effect on the requirements gathering process. For example, although the POC is primarily driven by EMAP scientists, some policy makers, and other scientists are involved in the process.

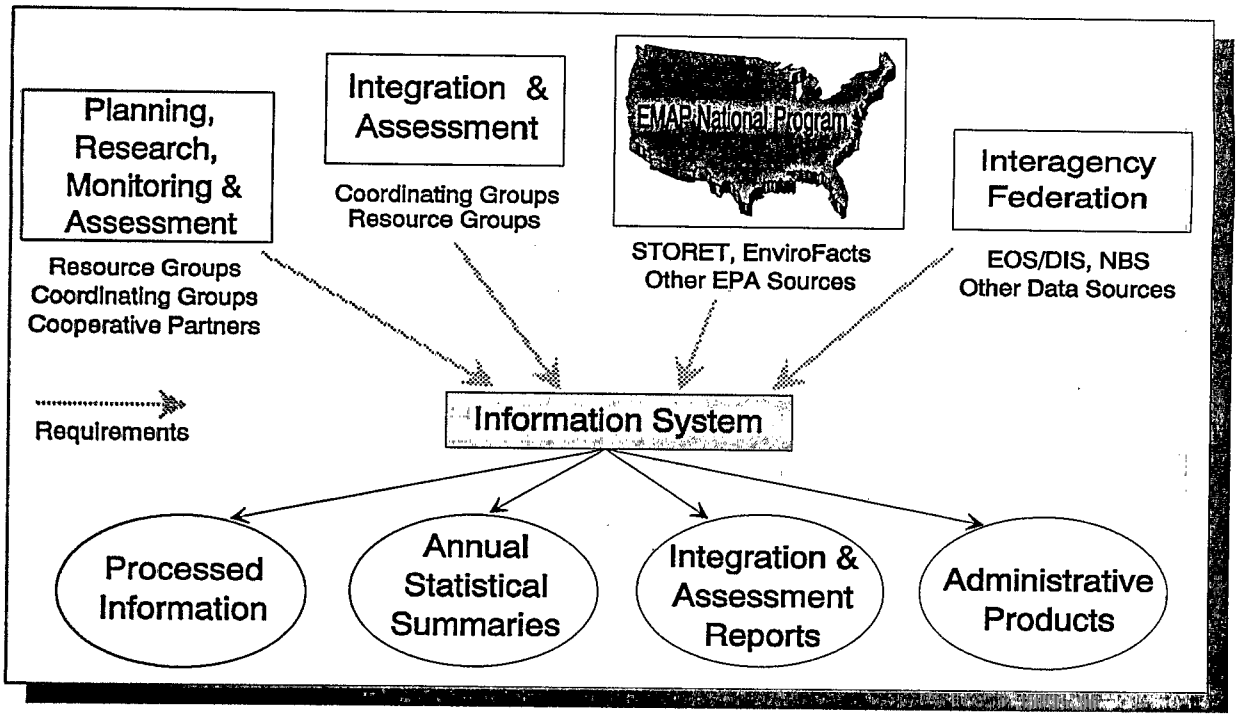

Figure ES-5. EMAP enterprise products

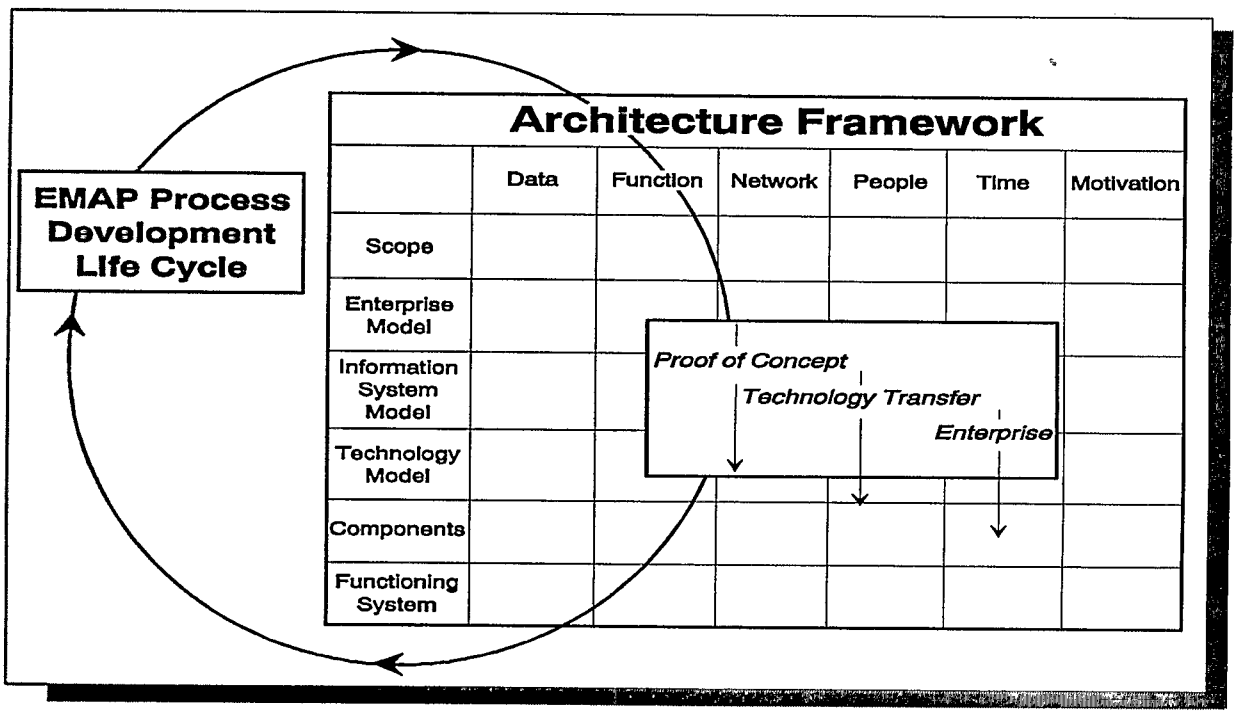

Figure ES-6. EMAP architecture framework.

EMAP Products

EMAP will provide three products that the EMAP IM system must support during the Plan period:

- Databases and derived information;
- Annual statistical summaries; and
- Integration and assessment reports.

Additionally, EMAP IM is designed to provide for administrative products that will be important to support EMAP as a National Program within EPA (Figure ES-5).

EMAP IM Technical Guidelines

The technical approach used to define EMAP IM system development and implementation will be based on these guidelines:

- Use existing National infrastructure and standards whenever possible;
- Concentrate on integration engineering rather than new system development;
- Use a systems engineering approach based on a formal design process incorporating functional pilots and prototypes; and
- Use emerging technology whenever possible in meeting user needs.

Evolutionary Approach

EMAP IM is designed to use an "evolutionary" systems engineering approach that adapts to emerging user needs and technology. This approach will permit ongoing use of operational systems while the development process continues. The distinguishing difference between the EMAP evolutionary approach and more traditional system engineering approaches is EMAP's greater involvement of users throughout the system development life cycle (SDLC). The SDLC elements include *concept, analysis, design, development, test,* and *operation.* Activities during each element of the SDLC will take into account current and projected user needs and technology, and incorporate user feedback and new technology into the system development process.

EMAP IM Framework

To assist in system development, EMAP IM uses the Zachman Framework[3] as an architectural guide during the Enterprise process. This Framework allows for an orderly approach for designing and building systems. It provides system developers a complete template against which to map sequential and/or parallel steps for carrying out development of technical and management solutions.

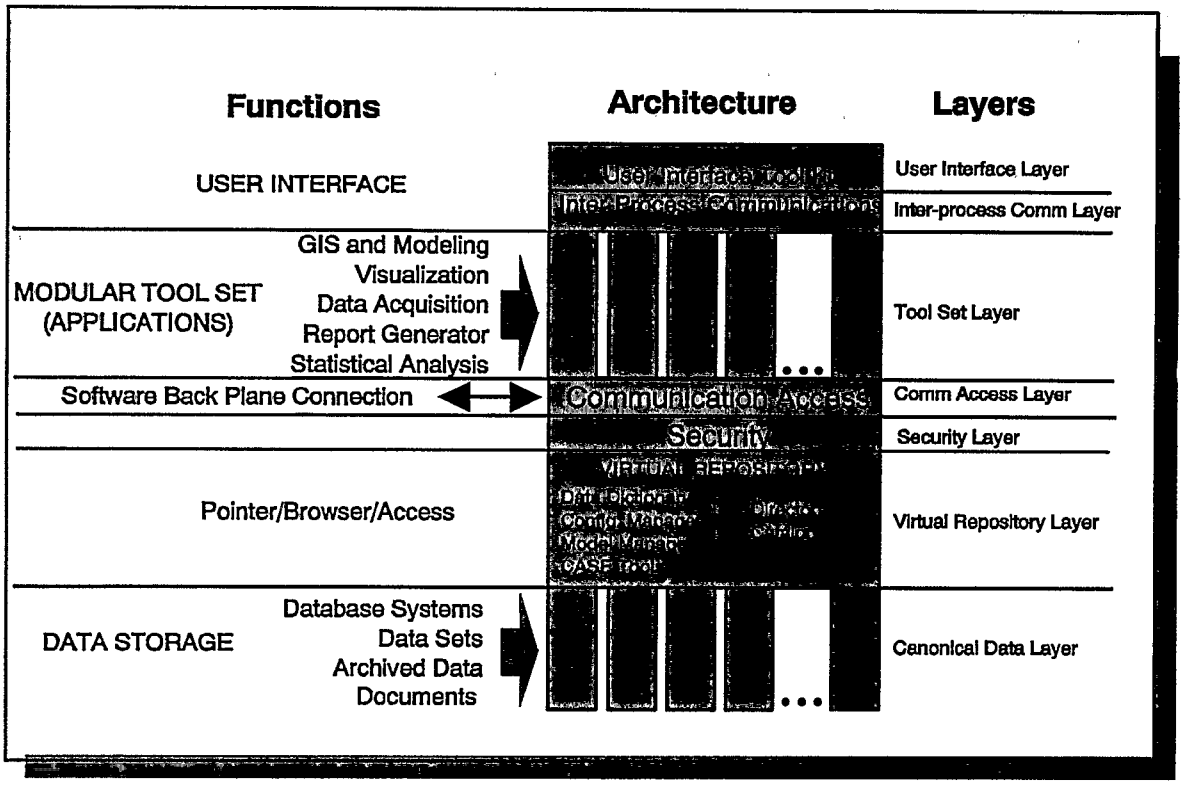

Figure ES-7. EMAP IM architecture.

The Zachman Framework stresses that the information system development process requires several architectures. Figure ES-6 depicts the Framework architectures: *Data, Function, Network, People, Time,* and *Motivation.* The levels of each architecture in the Framework are: *Scope, Enterprise Model, Information System Model, Technology Model, Components* and *Functioning System.*

Each Framework architecture (e.g., data) develops as the process progresses though the framework's life cycle from *Scope* through *Functioning System.* This process development life cycle is applied to each architecture within the Framework. This iterative process will be applied to each subsequent Enterprise component during the three IM implementation processes: *Proof-of-Concept (POC), Technology Transfer,* and *Enterprise Implementation.*

Using such a checklist ensures that all aspects of the system are addressed from an "enterprise approach" and that all participants in the development and implementation of the Enterprise have a common reference. (This Framework has been adopted for use by the EPA Office of Information Resources Management [OIRM] as a basis for all EPA systems.) The Framework also provides a technology-independent model to accommodate emerging hardware development and software changes.

EMAP IM Architecture

Another view of the EMAP architecture is provided by the seven layer model depicted in Figure ES-7. (It should be noted that these architecture functions will be provided, to varying degrees, by dispersed systems at multiple nodes.) Forming the foundation of the architecture is the canonical data layer. It is comprised of databases and data sets that can be accessed by tools from the tool set layer.

Above the database layer is the Virtual Repository layer. The components of the Virtual Repository are model management tools, Computer Aided Software Engineering (CASE) tools, configuration management tools, the data dictionary, the directory, and catalog. The first three tools mentioned in this layer and the data dictionary pertain to computer science metadata that describes data stored in a relational database structure. The directory and catalog components of the Virtual Repository layer contain scientific metadata that describe data sets and documents.

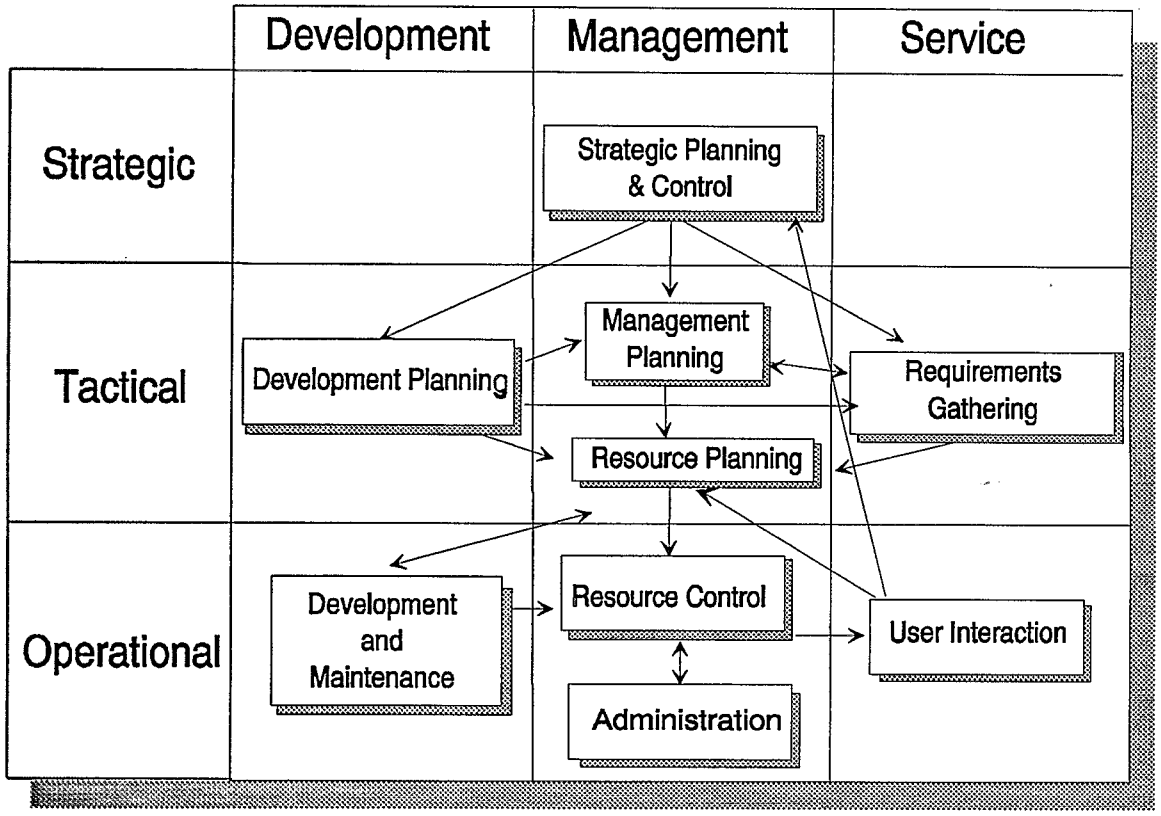

Figure ES-8. EMAP IM structure.

	Development	Management	Service
Strategic		• IM Coordinator • Coordinating Groups • Science Direction	
Tactical	• User Interaction & Planning • Resource Groups	• IM Coordinator • System Architecture • Technology Evaluation • Coordinating Groups • Resource Groups	• User Interaction & Planning • Interagency Data Interchange • Resource Groups
Operational	• GIS • Systems Engineering • System Support & Operations • Coordinating Groups • Resource Groups	• IM Coordinator • Coordinating Groups • Resource Groups	• User Interaction & Planning • Coordinating Groups • Resource Groups

Figure ES-9. EMAP IM functions.

The security layer provides a means for EMAP to ensure the long-term integrity of its collection by controlling access. Security procedures, in practice, can apply at the user interface layer, the communications access layer, and the canonical data layer.

The communications and user access layer represents the means by which data may be entered and retrieved from the data layer using the Virtual Repository. Note that the software back plane connects at this level and that access is provided to users of the back plane through the Virtual Repository. A user entering through the tool set layer has the option of directory or catalog access, or access through the flexible set of tools that are provided to the user community to facilitate their use of the data. The inter-process communication layer allows the individual tools to communicate by passing data from one to another for processing. Finally, the user interface layer provides easy invocation of the various tools available as part of the EMAP suite, and other tools available on the client platform.

EMAP IM Management

The EMAP Director has overall program responsibility. The Director of EMAP Center directs eight Coordinating Groups and is responsible to the EMAP Director. The Coordinator of Information Management, who heads one of the eight Coordinating Groups, is responsible for EMAP IM. The IM Coordinator relies on members of the eight Resource Group IM teams for coordination and implementation of EMAP IM.

EMAP IM is using an adaptation of an IBM Corporation information systems management process model[4] as the basis for project management as shown by Figure ES-8. This model has strategic, tactical, and operational levels which are sectioned into development, management, and service missions. The evolutionary approach of EMAP IM engages all levels and missions virtually simultaneously since operational implementation must continue to occur within the Resource Groups while tactical and strategic events are evolving.

The EMAP IM functional organization consists of teams focused on: systems architecture, science direction, user interaction and planning, interagency data interchange, advanced technology evaluation, systems engineering, and systems support and operations. Each of these teams has specific roles within the management structure (Figure ES-9). Additionally, ongoing coordination is required with Federal and EPA information resources management programs and offices.

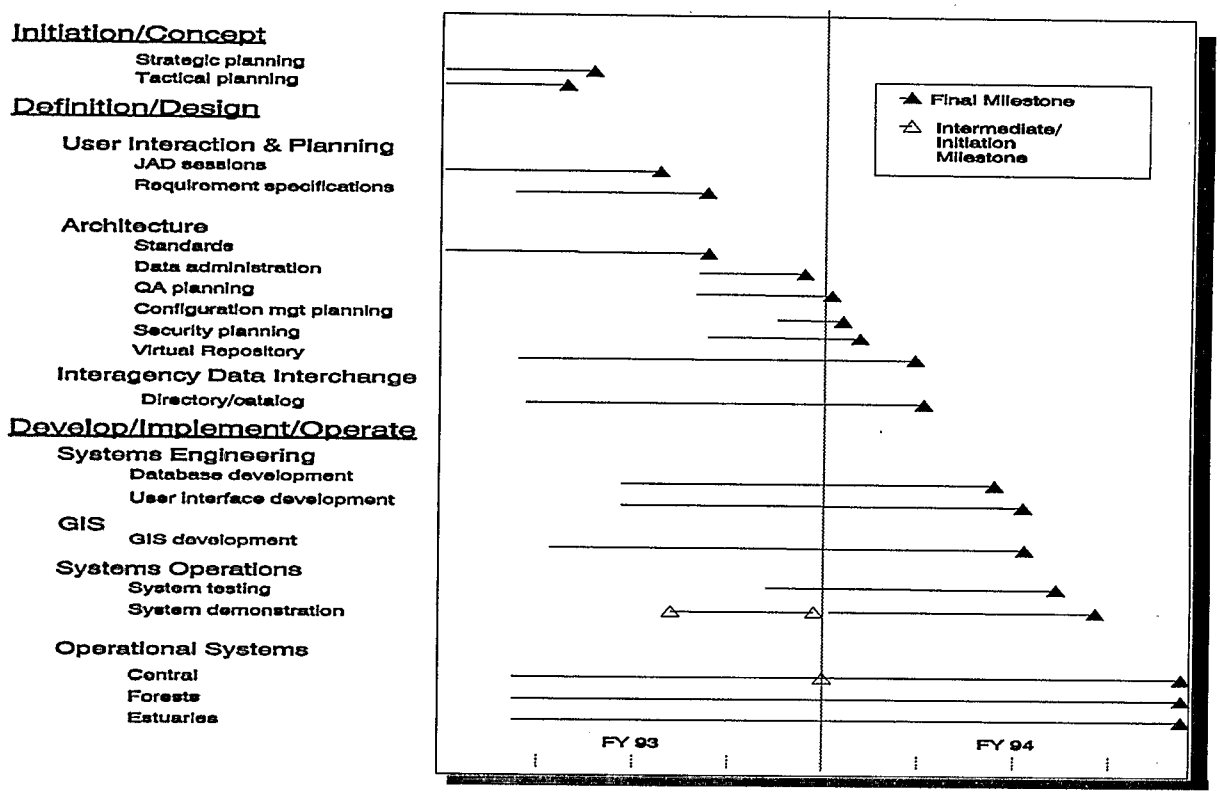

Figure ES-10. EMAP IM implementation schedule – POC.

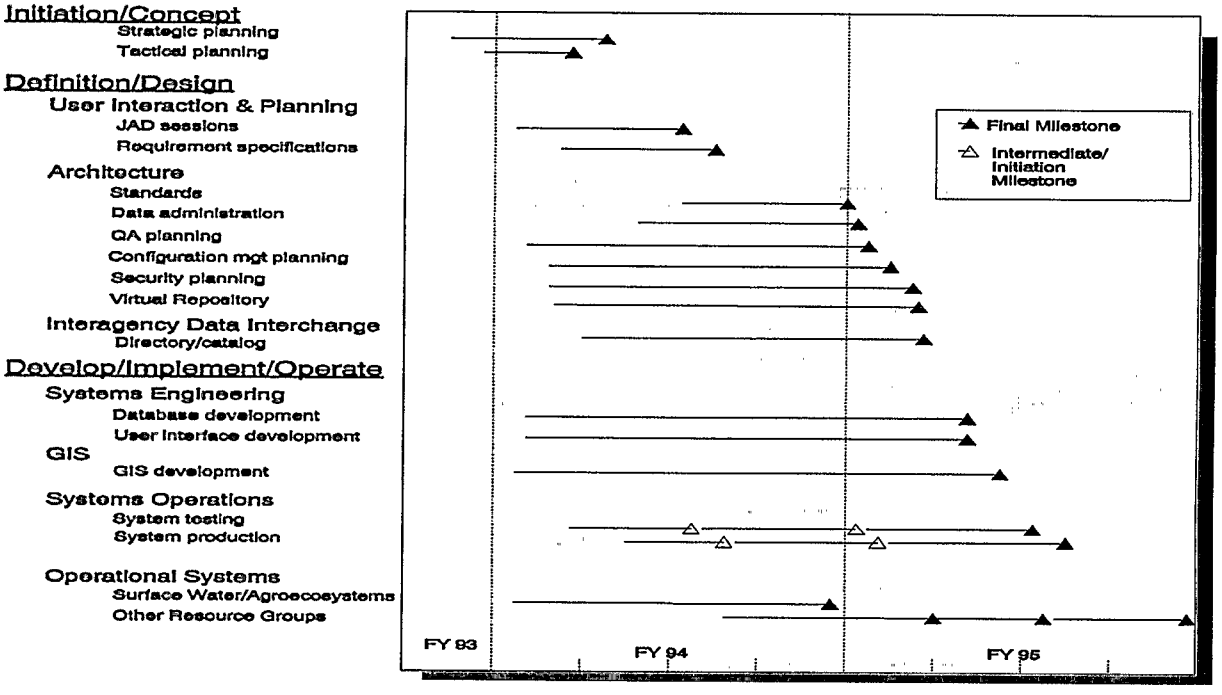

Figure ES-11. EMAP IM implementation schedule – Technology Transfer.

EMAP IM Implementation Processes

EMAP IM development and implementation consists of three IM process phases. First is the Proof-of-Concept (POC) process that identifies and applies standards, policies, tools, and procedures to the Forests and Estuaries Resource Groups and limited Coordinating Groups. The technology transfer phase consists of extending the environment validated during the POC to the remaining Resource Groups and an increasing number of Coordinating Groups. The third process is Enterprise Implementation, which engages other EPA and non-EPA programs in the systems development process. This phase re-engineers preceding implementation activities from an enterprise perspective.

The implementation schedules for each process are shown at Figures ES-10, ES-11, and ES-12. The schedules are arranged by program function into general system development life cycle (SDLC) categories titled Initiation/Concept, Definition & Design, and Implementation/Operation. Each of these categories then lists IM functional areas and planned deliverables. It is important to note that each step of the implementation process includes participation from other user categories. For example, all Resource and Coordinating Groups have been observing the *Planning, Research, Monitoring, and Assessment* POC in order to facilitate their role in the Technology Transfer process.

EMAP Schedule (1993-1997)

The schedules discussed below are based on the assumption that required resources will be available as depicted at Figure ES-13. If, however, constraints are placed on EMAP IM resources, the technical and management approaches specified in the Plan are sufficiently flexible to permit implementation to occur, but over a longer period of time. Also, it should be noted that capabilities scheduled for completion in the "out-years" are already in development to varying degrees. For example, the Virtual Repository, Geographic Information Systems (GIS), and user interface capabilities commenced with the 1993 POC.

1993 – *Research, Monitoring, and Assessment* POC. This demonstration was designed to enable two Resource Groups to collect, manage, and share distributed data.

1994 – *Planning, Research, Monitoring, and Assessment* Technology Transfer. This process will share the activities and technology from the POC with the other Resource Groups and Coordinating Groups. The system will: include the initial data dictionary, perform basic analysis, provide initial user interface, provide GIS and initial report generation tools, and perform

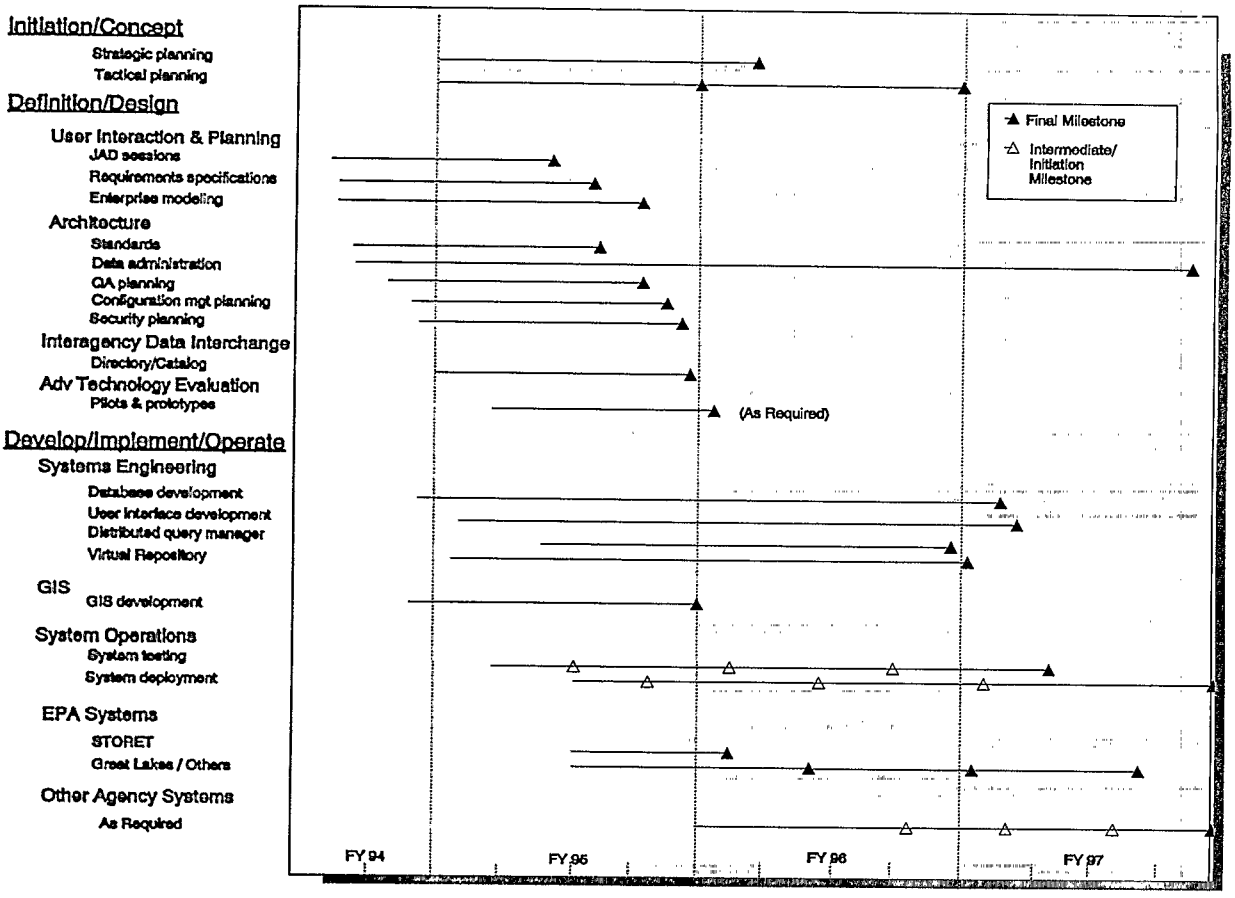

Figure ES-12. EMAP IM implementation schedule – Enterprise.

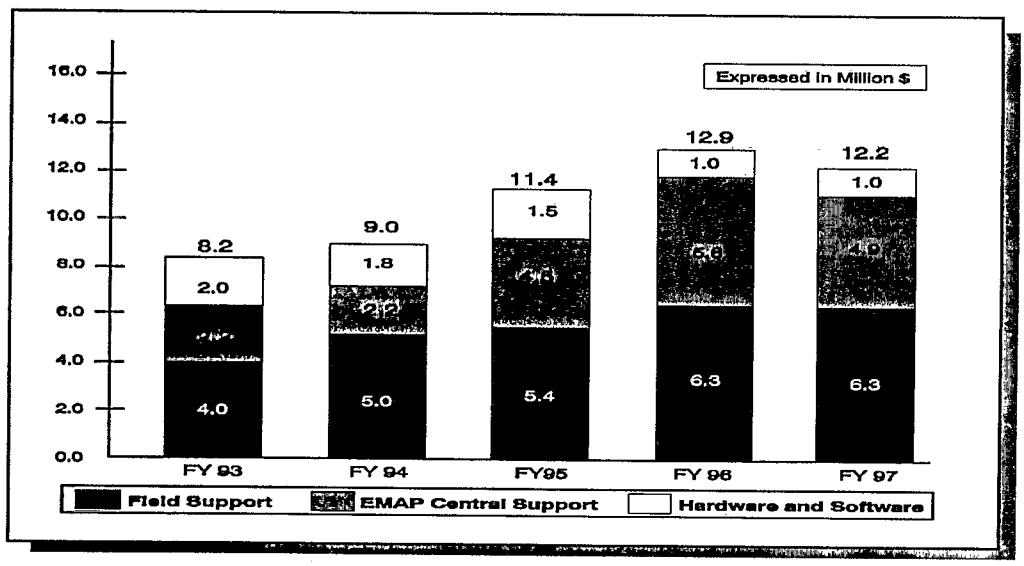

Figure ES-13. EMAP IM required budget.

internal data exchange adhering to established OIRM data standards.

1995 – *Integration and Assessment* Enterprise Implementation. The system should provide: improved data verification, integration of improved data acquisition tools, a fully capable model manager; enhanced user interface with visualization, limited reporting across EMAP's ecological resources, survey analysis methods and interfaces, and internal data transformation methods.

1996 – *EMAP National Program within EPA* Enterprise Implementation. The system will provide interoperability with other EPA programs. This should include enhanced data acquisition tools, accessible metadata and methods using a Virtual Repository, enhanced analysis and aggregation methods, a common user interface across many EPA platforms, an enhanced user interface featuring multi-media, and mechanisms for initial intra-agency data exchange.

1997 – *EMAP Interagency Federation* Enterprise Implementation. System interoperability among EMAP, EPA and other agencies should provide external data exchange.

EMAP IM Budget (1993 - 1997)

The budget estimates provided in this Plan include requirements for Resource Group field support teams, support for the IM Coordinator and EMAP Central, and hardware and software costs. Using FY 93 as a baseline, the figures for FY 94 - FY 97 include resources anticipated for EMAP from all supporting Cooperative Partners and other Federal agencies.

Summary

The scope of this Plan extends from FY 93 through FY 97. However, EMAP IM systems will continue to operate beyond 1997 as the user base expands. These users will impose increasingly demanding requirements that will become more pronounced as EMAP enters the *EMAP Interagency Federation.*

With the assimilation of increasing data and information from other national and international ecology-related programs, EMAP IM is designed to become an "information highway." It will serve as a powerful instrument for aggregating key environmental information to assist policy makers and managers at all levels in affecting the positive changes necessary to preserve the global environment.

Section 1

INTRODUCTION

1.1 Purpose of the EMAP IM Strategic Plan

This Information Management Strategic Plan outlines implementation of an information management approach to fulfill the evolving information management (IM) needs of the Environmental Monitoring and Assessment Program (EMAP). The objectives of this Plan are to provide U.S. Environmental Protection Agency (EPA) management with resource estimates; EPA environmental scientists and collaborators (other agencies, universities, and industry) with a common set of user expectations; and EPA information managers with a systems engineering road map for use in directing activities over a 5 year period.

1.2 Scope of the EMAP IM Strategic Plan

This Plan covers a 5-year planning period. Correspondingly, references in this Plan focus principally on the users and information systems technology that are applicable through 1997. Specifically, this Plan defines the following aspects of EMAP IM systems:

- A vision of EMAP and how EMAP IM supports that vision;
- EMAP objectives and how EMAP IM supports those objectives;
- The process for addressing the needs of the environmental science community;
- Who is expected to use EMAP IM and in what priority;
- Guiding principles for EMAP information management;
- A technical approach that will evolve over time in response to changing technologies and EMAP user objectives;
- A project management model suitable for a systems engineering effort of this scope; and
- An estimate of the resources required to be successful.

Certain technologies and processes that are not included in the Plan eventually will be part of EMAP IM. These topics will be addressed in separate "white papers." These papers will be an integral part of the conceptual or analytical components of the system development life cycle (SDLC) for EMAP IM, and will be germane to the development and implementation of the EMAP IM Enterprise. These papers will be prepared during FY 1994 and FY 1995 for inclusion in the ongoing strategic and tactical planning processes. The papers include the following topics:

- Database management systems (DBMS) evolution from relational -- to object-oriented databases;
- Enterprise architecture conceptual and logical design;
- Enterprise architecture back plane compatibility with software tools;
- Data and process model integration;
- DBMS and data set access through a Virtual Repository;
- Interagency data/information exchange; and
- Performance documentation requirements for data access, data entry, and human factors issues.

1.3 Intended Audience for the EMAP IM Strategic Plan

The intended audience and users of this Strategic Plan include:

- EPA and EMAP management -- to support budgetary allotments, interagency cooperation, and overall direction;
- EMAP scientists -- to provide uniform expectations for EMAP IM;
- EMAP information managers and staff -- to provide a common development and integration view;
- Environmental scientists -- to illustrate how they can participate in the project;
- Scientific reviewers -- to provide an opportunity for input and to support EMAP Program credibility; and
- Other governmental agencies -- to enable cooperation in related environmental programs.

1.4 Information Management Life Cycle Approach

A typical information management life cycle consists of five phases: 1) initiation; 2) concept; 3) definition and design; 4) development and implementation; and 5) operation. These life cycle phases, as they generally correlate to life cycle products, are shown in Table 1.4.1.

EMAP IM Life Cycle	Typical Life Cycle Products
Initiation	• Initiation Decision Paper • Project Management Plan • Data Automation Plan
Concept	• Strategic Plan • System Concept Paper • System Test Document • Acceptance Test Document • Data Management Plan • Concept Decision Paper
Definition/Design	• Architecture Standards • Configuration Management, Accounting, and Change Control Records • Detailed Functional Requirements • Detailed Data Requirements • Definition Decision Paper • Requirements Data Dictionary • System Design Document • Design Decision Paper • Physical Database Design • Design Data Dictionary
Development/ Implementation	• Development System Software • Development Databases • Maintenance Manual • User Manuals • Operation Manual • Security Manual • User Support Materials • Development Decision Paper • Production System Software • Production Databases • Production Data Dictionary • Implementation Decision Paper • Training Report
Operation	• Performance Report • Post Implementation Evaluation Report • System Evaluation Report • System Disposition Report • Achieved/Incorporated Data • Achieved/Incorporated Software • Achieved Life Cycle Products

Table 1.4.1. EMAP IM Life Cycle with Typical Life Cycle Products

EMAP IM adheres to these IM Life Cycle phases but within the combined context of the Zachman Framework[3] (as explained in Section 5), EPA's Office of Solid Waste and Emergency Response (OSWER) guidance[5], and Rapid Application Methodology[6] (RAD). This relationship is shown in Table 1.4.2.

Many specifications and plans will be developed and refined during the EMAP IM Life Cycle process.[7] The Life Cycle products from Table 1.4.1 that are applicable to EMAP IM will be specified in the EMAP IM Tactical Plan. This EMAP IM Strategic Plan is a key component to ensuring continuity throughout an evolutionary life cycle process.

Zachman Framework	OSWER Guidance	RAD Methodology
Business Scope	Initiation Phase	Requirements Planning Phase
Business Model	Concept Phase	
Information Systems Model	Definition Stage	User Design Phase
Technology Model	Design Stage	User Design Phase
Technology Definition	Development Stage	Construction Phase
The Information System	Implementation Stage	Cutover Phase

Table 1.4.2. Relationship between Zachman Framework, OSWER Guidance and RAD Methodology

Section 2

BACKGROUND

To understand the evolving needs for information management within the Environmental Monitoring and Assessment Program (EMAP), this section provides a brief description of EMAP. This is followed by a description of information management (IM) within EMAP, including the definition and vision of EMAP IM and its objectives. The EMAP organization is also discussed in Appendix A with relationships and responsibilities to other EPA and Federal organizations.

2.1 EMAP Vision

The Environmental Monitoring and Assessment Program is responding to the growing demand for information about the changing condition of our ecological resources.[8] Once EMAP has been fully implemented, it will help answer many questions, including the following:

- What is the current extent, geographic distribution, and condition of our ecological resources, including estuaries, lakes, streams, wetlands, forests, grasslands, deserts, and agricultural areas?
- What portion of these systems is degrading or improving, in what areas, and at what rate?
- What are some possible causes of adverse condition in these systems?
- Are adversely affected systems responding as expected to control and mitigation programs?[9]

To answer these questions, EMAP will collect and analyze data from many of the Nation's ecological resources. Data collection will be based on ecological and biological indicators that characterize exposure, stress, and response of the ecological

resources to environmental changes, either natural or man-made.[10] An ecological indicator is defined as any characteristic of the environment that can provide quantitative information on the condition of ecological resources, magnitude of stress, exposure of a biological component to stress, or the amount of change in conditions.

EMAP field monitoring studies are designed to provide the majority of the data. Field monitoring studies are based on probability sampling to allow characterization and assessment on a regional and national scale. Historical data, field data, and remotely sensed data collected by other Federal agencies and programs supplement EMAP field monitoring activities.

2.1.1 EMAP Goals and Objectives

EMAP goals are to monitor and assess the condition of the Nation's ecological resources and to contribute to decisions on environmental protection and management. To accomplish these goals, EMAP works to attain four objectives:[1]

1. **Estimate the current status, trends, and changes in selected indicators of the condition of the Nation's ecological resources on a regional basis with known confidence;**
2. **Estimate the geographic coverage and extent of the Nation's ecological resources with known confidence;**
3. **Seek associations between selected indicators of natural and anthropogenic stresses and indicators of condition of ecological resources; and**
4. **Provide annual statistical summaries and periodic assessments of the Nation's ecological resources.**

These objectives support EMAP's goals and seek to provide scientific information useful to decision makers. In turn, decisions regarding environmental protection and management require that the important societal value associated with our ecological resources are identified and stated clearly.

2.2 Information Management (IM) within EMAP

Information management within EMAP is a common integrating process required to support attainment of EMAP objectives.[1] The EMAP IM infrastructure will use resources (i.e., personnel, hardware or software) at geographically dispersed locations. The EMAP IM process will develop or support information systems based on the inputs and needs of several major groups of users within, and outside of, EMAP. The EMAP program, from its inception, has emphasized access to data at the agency level as well as the program level. Correspondingly, EMAP IM will provide access to other government agencies' data through EPA-networked systems.[11]

2.2.1 EMAP IM Vision

Information management as it applies to EMAP requires comprehensive, integrated systems that facilitate communication of environmental data from diverse sources spatially (over distance) and temporally (over time) in a form that users can easily access. An IM framework will manage this data through established protocols, procedures, and standards. Simply stated, the vision for EMAP IM is:

> ***To provide information management products to support monitoring and assessment of the condition of the Nation's ecological resources. Information management will provide the infrastructure for turning raw scientific measurements into information that can improve the quality of our environment.***[12]

The IM component of EMAP is the vehicle through which the total EMAP program manages information from collection to distribution (as information products) and archival. EMAP IM systems should deliver information of known quality to users quickly and easily, in a form usable for the function that each user is performing.

All systems are designed to meet National standards that facilitate change and growth. The Ecological Resource Group users are involved on an iterative basis in the design and development of the IM systems and the resulting products. Care is given to the unique spatial and temporal diversity of the EMAP program ensuring that data and methods for its collection, analysis, and reporting will evolve over time with no loss of information. Data and metadata must be documented so that current results can be replicated in 20 years using the same data or information. Four broad categories of data are managed by EMAP:

- Monitoring data with a spatial and temporal component collected by EMAP Resource Groups;
- Descriptive information to completely document the metadata of all data collected;
- Derived data and its description; and
- Descriptive and supportive data collected from sources external to EMAP.

Also, EMAP IM systems will manage internal information that supports the data and its use or analysis. These support data include program plans, algorithms, procedural handbooks, publications, contact lists, and others.

The elements of the IM infrastructure are a variety of electronic and manual technologies including, but not limited to:

- Data acquisition for automated field data collection and inventory management;
- Data management to address data collection, processing, storage, cataloging, documentation, and security issues;

- Data analysis to allow the user to investigate spatially and temporally diverse data for one or more ecological resources through interprocess communications and database interoperability;
- Information product generation to allow production of databases, graphics, reports; and
- Communications to aid in distributing results of data analysis to a variety of audiences through network and data exchange protocols.

2.2.2 EMAP IM Objectives

The EMAP IM objectives are linked to the EMAP program objectives specified in Section 2.2.1. EMAP IM intends to provide IM support to the Resource Groups in their planning, research, monitoring, and analysis efforts so that differences in the IM environment of the groups are minimized. Further, EMAP IM is building an information management infrastructure that supports the total EMAP program with integrated data management, interfaces, standards, policies, and procedures. EMAP IM objectives are outlined in Table 2.2.2.1.

EMAP Program Objectives	EMAP IM Objectives
Estimate the current status, trends, and changes in selected indicators of the condition of the Nation's ecological resources on a regional basis with known confidence.	• Assist in structuring, developing, maintaining, operating, and/or deploying: - Resource Group databases for planning, research, monitoring, and analysis. - Access mechanisms to the databases. - Manipulation mechanisms (algorithms). - Display mechanisms required for planning, research, monitoring, and analysis.
Estimate the geographic coverage and extent of the Nation's ecological resources with known confidence.	• Assist in structuring, developing maintaining, operating, or deploying: - Access mechanism to EMAP Geographic Reference Data (GRD). - Manipulation mechanisms (algorithms). - Display mechanisms required to analyze this information.
Seek associations between selected indicators of natural and anthropogenic stresses and indicators of condition of ecological resources.	• Ensure a distributed database structure that allows integration of information internally, and with external sources allowing responsibility for the data to reside with the "owners." • Assist in the development, maintenance, operation, and deployment of cross-cutting tools.

Table 2.2.2.1. EMAP IM Objectives

EMAP Program Objectives	EMAP IM Objectives
Provide annual statistical summaries and periodic assessments of the Nation's ecological resources.	• Ensure that all data, metadata, and information required to meet the objective are available, in some cases by leveraging the central node of the database system. • Assist in the development, maintenance, operation, and deployment of access mechanisms and display mechanisms useful in meeting this objective.

Table 2.2.2.1. (continued)

2.2.3 EMAP IM Products

The specific goals or products that will result from the information management objectives and infrastructure, as shown in Figure 2.2.3.1, are:

- Processed and aggregated data and information derived from scientific, external, and corporate databases;
- Annual statistical summaries;
- Integration and assessments reports; and
- Administrative products.

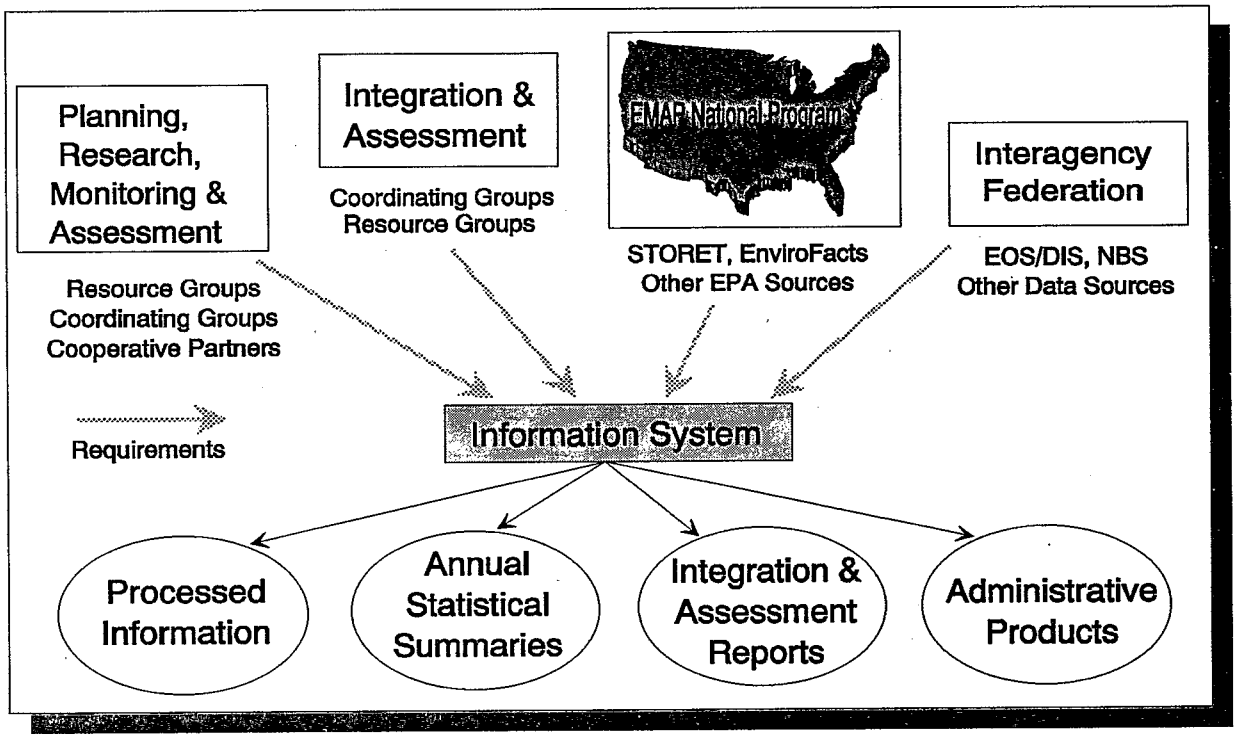

Figure 2.2.3.1. EMAP enterprise products.

2.3 EMAP Organization

The Environmental Monitoring and Assessment Program (EMAP) is an interdisciplinary, interagency program being designed and initiated through the U.S. Environmental Protection Agency's (EPA) Office of Research and Development (ORD). The program's objectives require that EMAP be an interagency program in which EPA is but one of the participants. Within ORD, EMAP is organized as depicted in Figure 2.3.1. EMAP management within EPA, as of December 3, 1993, is comprised of:

- EMAP Director;
- Two Deputy Directors and one Associate Director;
- EMAP Center Director;
- EMAP Headquarters Associate Director; and
- Eight Resource Group Technical Directors.

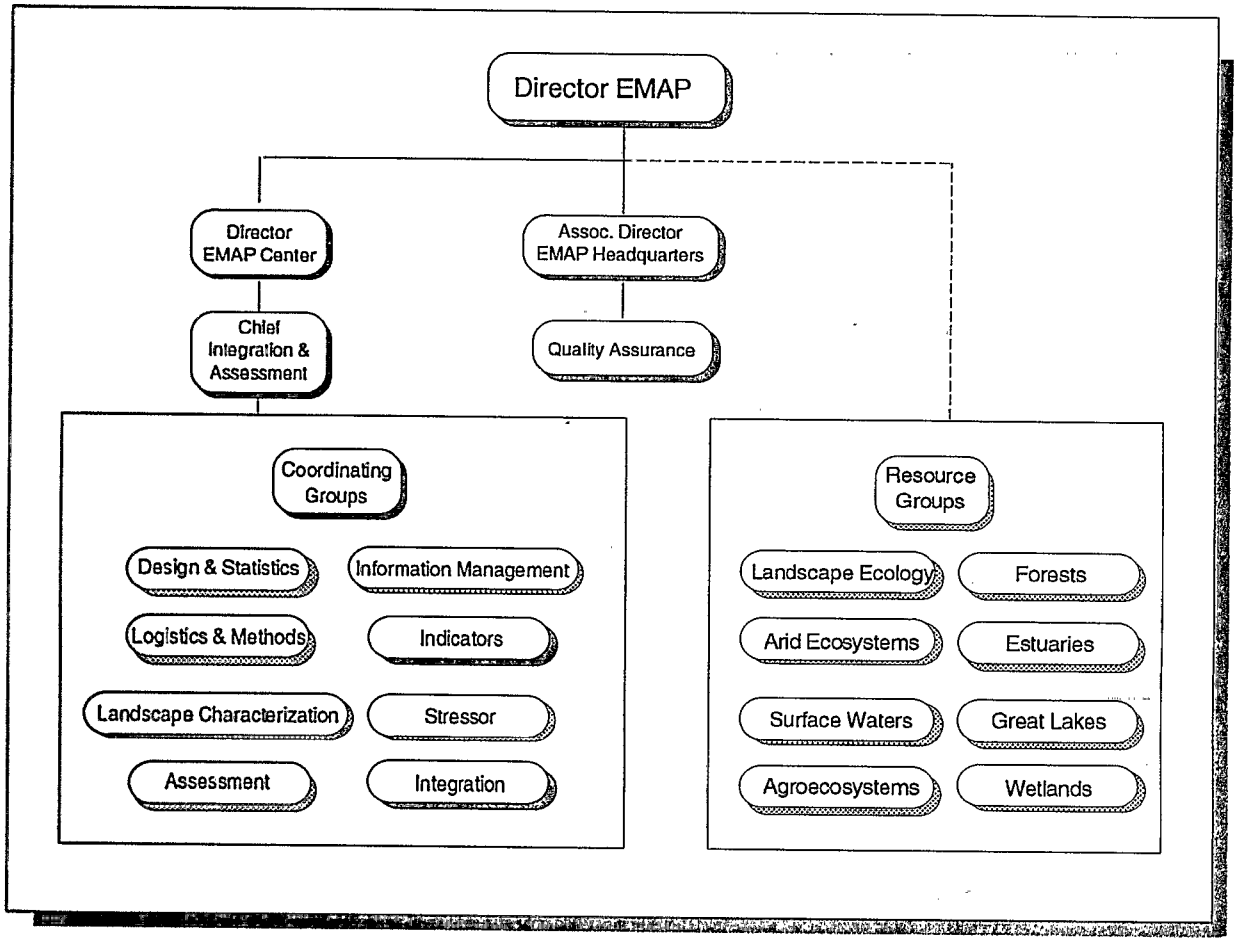

Figure 2.3.1. EMAP organizational structure.

EMAP has eight ***Resource Groups*** that focus on ecological resources: Great Lakes, Estuaries, Arid Ecosystems, Agroecosystems, Forests, Surface Waters, Wetlands, and Landscape Ecology. These Resource Groups are geographically distributed across the U.S. and often conduct joint activities with such Cooperative Partner organizations as the National Oceanic and Atmospheric Administration (NOAA), the U.S. Forest Service (USFS), the U.S. Department of Agriculture (USDA), and others. This cooperation includes sharing of EPA's and other agencies' computing resources.

The eight ***Coordinating Groups*** are: Design and Statistics, Information Management, Logistics and Methods, Indicator Development, Landscape Characterization, Stressor, Assessment, and Integration. The Quality Assurance position, which is essential for the success of EMAP IM, reports to the EMAP Headquarters Associate Director but also performs a Coordinating Group function. These groups perform functions that cut across EMAP Resource Groups. And, since EMAP is an interagency program, many of its activities are distributed nationally.

2.3.1 EMAP IM Responsibilities

The EMAP IM representatives from Resource and Coordinating Groups support the IM Coordinator for the implementation of EMAP IM. Members of this IM group are responsible for the implementation of this Plan. The EMAP IM representative from each Resource Group is responsible for providing hardware, software, documentation, and systems support for that Resource Group. The IM Coordinator assists the Resource Groups with their EMAP IM requirements; but, is directly responsible for providing systems support for the Coordinating Groups. The combined IM Group provides information standards, technical procedures, and IM guidance to EMAP.

EMAP IM systems must provide broad support to:

- EMAP senior management in planning, coordinating, and implementing activities to evaluate environmental status and trends;
- All Coordinating Groups conducting information management activities, including collection, transfer, analysis, and presentation of EMAP data;
- EMAP production of standard reports;
- Such EMAP internal collaborators as EPA's Office of Information Resource Management (OIRM);
- Such EMAP/EPA external collaborators as NOAA through the applicable Resource Groups;
- A variety of users external to EMAP: EPA regional offices, States, legislative decision makers, the scientific community, other agencies and institutions, and the public and private sectors; and
- Assessment activities performed by the above groups.

EMAP IM systems are designed to meet continuously evolving user needs using appropriate technology advances in hardware, software, and telecommunications. EMAP IM supports both centralized and distributed elements of EMAP. Because EMAP is an evolving scientific program, EMAP IM systems are designed to incorporate such techniques as rapid prototyping, open systems design, continuous user involvement, and other techniques under consideration by EPA IM working groups. Resource and Coordinating Groups can also propose standards that influence development. They can evaluate prototypes, hardware, software, and other IM alternatives, and provide the initial needs' assessment for future systems.

Interaction among the EMAP IM participants is shown at Figure 2.3.1.1. Note that each Resource Group has its own IM infrastructure. EMAP IM assists these Resource Groups with systems development from an EMAP-wide perspective. The Coordinating Groups, however, do not possess such IM capabilities and EMAP IM provides direct support for their systems development and operations.

In summary, these responsibilities contribute to a set of principles to guide EMAP IM systems development. EMAP IM is responsible for implementing this Plan within Federal and EPA Information Resources Management (IRM) guidelines and meeting the scientific communities' expectations for EMAP.

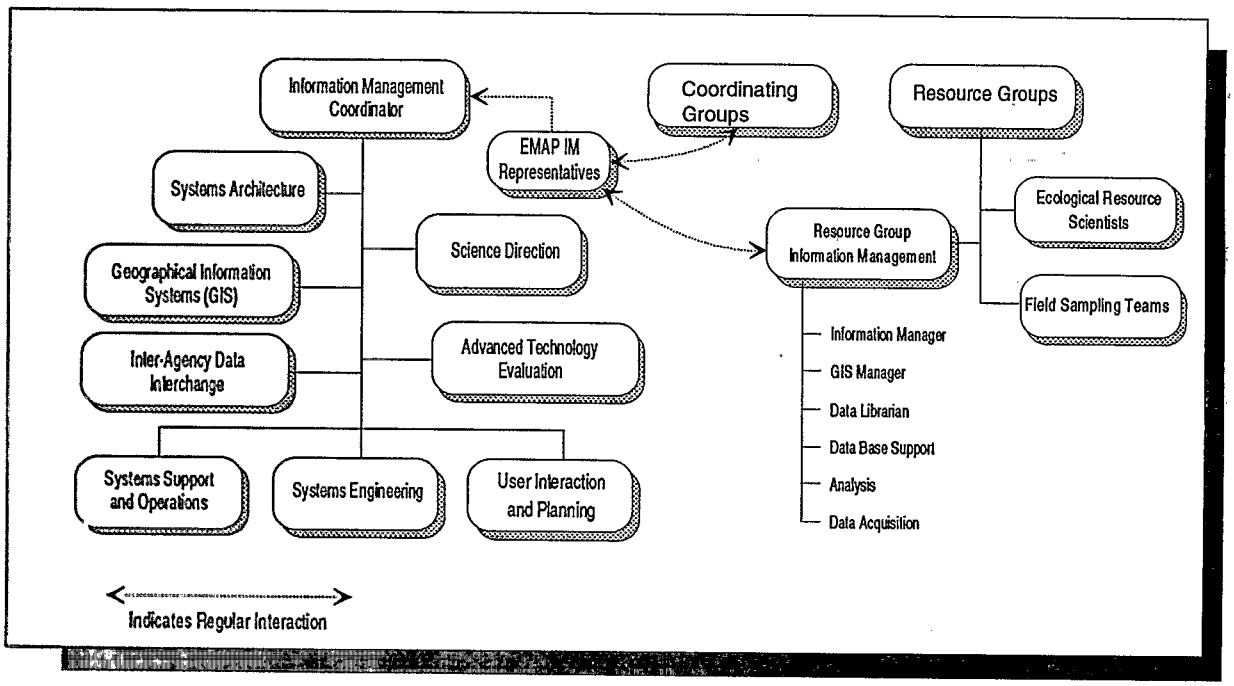

Figure 2.3.1.1. EMAP IM interaction with Coordinating and Resource Groups.

Section 3

EMAP USERS AND REQUIREMENTS

EMAP IM systems must provide data to a diverse set of users in a variety of formats. Users range from sophisticated scientific users who extract crosscutting data sets for detailed quantitative analyses to users from the general public who require information products. It is intended that EMAP IM principal users be the driving force in defining requirements and then validating the development of the "system." This section describes the categories of users, their current roles in systems development, and their information and data requirements.

EMAP is a multi-agency program that is heavily dependent on Cooperative Partners. The original Cooperative Partners were brought into EMAP largely through the efforts of individual Resource Groups who had common interests and interactions with other agencies. Examples of Cooperative Partners are National Oceanic and Atmospheric Administration (NOAA), U.S. Forest Service (USFS), U.S. Soil Conservation Service, U.S. Fish and Wildlife Service, and the Bureau of Land Management.[1] These Cooperative Partners actively participate with EMAP Research Groups and with EMAP IM systems development.

Other participants in EMAP include the EPA Geographic Initiatives. EPA initiated the Chesapeake Bay, Great Lakes, and Gulf of Mexico geographic programs to make ecological assessments of specific areas. Data collected by the Geographic Initiatives are of interest to Resource Groups; therefore, EMAP IM must coordinate with these programs.

3.1 EMAP IM Users

For purposes of this Strategic Plan, there are two categories of EMAP data and information users: principal users that have immediate EMAP IM requirements, and other users whose IM requirements are not yet as clearly defined (as shown in Figure 3.1.1). The first category, Principal EMAP Users, includes three groups of participants: Resource Group users, Coordinating Group users, and such EPA users of EMAP data as the Office of Research and Development through their laboratories.

These initial users of EMAP data are responsible for collecting, managing, and analyzing data to the point of data validation for further use. This category also includes EMAP's Cooperative Partners in the data collection phase.

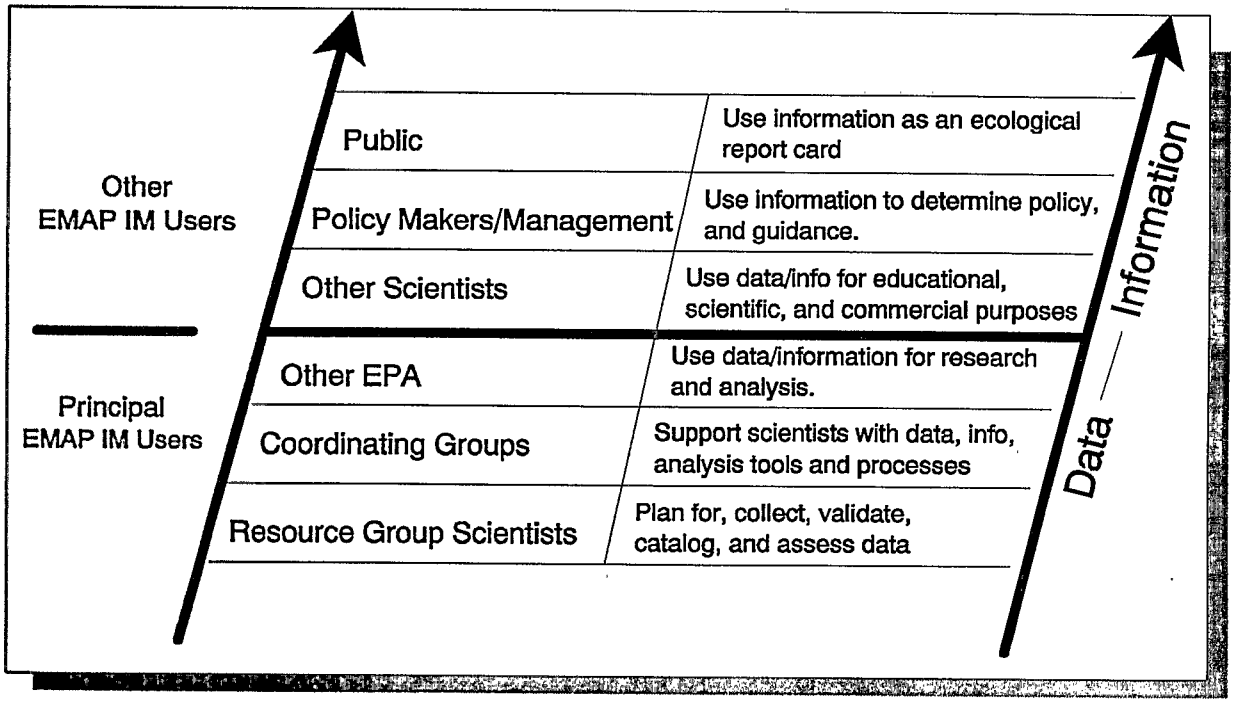

Figure 3.1.1. EMAP IM users.

3.1.1 Resource Group Users

Background

A typical user within a Resource Group has a scientific background but is not likely to be an expert in information technology. Supporting these scientific users at the Resource Group level are information managers, who store, manage and make data available to the Resource Group scientists. A heterogeneous variety of computing hardware, software, and networks is used throughout the Resource Groups.

Requirements

Resource Group scientists include Technical Directors, field scientists, analysts and others. Their primary activities are planning, implementing, analyzing and reporting. Planning involves designing the research program, developing ecosystem indicators, and conducting associated logistics work. Implementation involves the field measurement and data collection of information required to generate the ecosystem indicators. Once collected, data is assessed, analyzed, and used to produce various EMAP reports including the Resource Group's Annual Statistical Summary. A Resource Group information manager's primary function is to manage and provide access to Resource Group data. Primary requirements for Resource Group users include:

- **Planning/Design.** Data previously collected must be available for review prior to conducting new field operations. Especially during the research phase, there is a great demand for easy ad hoc access to existing data. Since EMAP uses a probability-based sampling design over time and space, pre-loading information into the database, (such as location and timing), is helpful in design planning prior to sampling and will speed the data entry process.
- **Sampling.** Sampling time may be reduced using automatic sampling devices and field computers to more rapidly sample and input data. During data collection activities, the data should be transferred into data management systems as quickly as possible.
- **Access.** Data collected should be available internally to the Resource Group as quickly as possible.
- **Data Integrity and Accuracy.** It must be possible to recreate accurate data. If data is not accurate, it could have wide-spread ramifications including misleading research efforts, incorrect estimation of trends, and possible misdirection of U.S. environmental policy.
- **Documentation.** EMAP data must be consistent over an extended period of time. Consequently, adequate documentation of the data and associated metadata needs to be collected, managed, and provided to users of the data.
- **Analysis and Reports.** Reports such as Annual Statistical Summaries must be generated to represent the environmental state of resources for which each group is responsible. These reports are generated from the data sets and databases created from sampled data, and should be automated as much as possible.

Use of EMAP IM in Meeting These Requirements

EMAP IM supports the Resource Groups in developing operational information management systems in a distributed environment. While each Resource Group IM system is developed and maintained by its respective Resource Group, they can leverage common EMAP data and information located in EMAP Central databases, as well as data and metadata located at other Resource Groups. EMAP IM assists

the Resource Groups by providing guidance and methods for electronically locating and accessing external data. Such assistance may involve technology evaluations of: global positioning systems for accurately determining sample locations; portable data recorders for automating field data collection; and high-speed distributed networks for uploading data to base stations or laboratories. In order to meet the EMAP objectives noted in Section 2, EMAP IM provides:

- Tools to be used in planning and design, such as a model manager, configuration manager, data dictionary, directory and catalog (elements of the Virtual Repository discussed in Chapter 5). These tools will provide information managers with a standard interface to Resource Group data and metadata, as well as other EMAP or external data and metadata;
- Assistance implementing interfaces between automatic sampling devices and field computers, and the final data management system;
- Directory and catalog tools to aid in accessing collected data sets;
- Standards for data integrity, security, and configuration management;
- Guidance on appropriate levels of metadata to accumulate as documentation during data collection activities; and
- Availability of external databases and access to other Resource Group databases to aid in analysis, and assistance in developing procedures to automatically generate reports from collected data.

3.1.2 Coordinating Group Users

Background

Like Resource Group scientific users, a typical user within a Coordinating Group has a scientific background but is not likely to be an expert in information technology. Coordinating groups do not have Information Management resources and rely on EMAP IM for implementing and maintaining their IM needs. These users have diverse computer and communications network environments.

Requirements

A typical Coordinating Group user is a Technical Coordinator who interacts with several, if not all, of the Resource Groups in a cross-cutting functional capacity. These users support specific Resource Group activities and facilitate the development of EMAP products and communication among other users of EMAP data. Primary IM requirements for Coordinating Group users include:

- **Support.** Design, program, implement and operate information management systems. This includes the development of databases and systems with which data and metadata can be accessed;
- **Availability.** Quick availability to Resource Group data;

- **Access.** Easy access to Corporate and Resource Group data;
- **Consistency.** Consistent format of common data between groups; standards for common data formats and usage guidelines are required for data aggregation; and
- **Data Quality.** Known quality of data that is sufficient for the intended use.

Use of EMAP IM in Meeting These Requirements

EMAP IM is providing system development and direct support for the Coordinating Groups, including database development, and integrated tool set development for locating, storing, and accessing data throughout the EMAP program. EMAP IM will provide:

- Direct development support of cross-cutting databases at the Central Node, such as the database for the Methods Coordinating Group, that describes all types of sampling and analysis methods used throughout EMAP; and guidance to all development groups to ensure that provisions are made in all databases to include needs of other Coordinating Groups, such as Design and Statistics;
- The infrastructure to access sampling data collected by all Resource Groups that are readily available and accessible;
- A Virtual Repository to index, summarize and provide detailed documentation of available data;
- Draft standards to provide consistency of data formats, data codes, and data exchange standards; and
- Guidance on developing data quality indicators that are consistent across EMAP and other EPA research programs.

3.1.3 Other EPA Users

Background

In addition to the EMAP Resource and Coordinating Groups, scientists within the EPA's Office of Research and Development (ORD) or other EPA Laboratories can use EMAP's information. For example, EPA regional offices may use EMAP data to assess the condition of environmental resources in individual regions. As in other groups, these users characteristically have a scientific background with a varying amount of experience with information technology. A heterogeneous variety of computing hardware and software will be used by this group.

Requirements

These users can apply EMAP data and information in their own research and analysis. Primary IM requirements include:

- Access to other EPA program information systems and data, such as the Great Lakes and Gulf of Mexico programs;
- Ease in locating EMAP data and reports – because these users will not be experts in EMAP data, the ability to identify and describe EMAP information will be crucial;
- Ability to identify metadata and other data documentation – required to determine if EMAP data is applicable to their specific research areas;
- Quality of data must be known and well documented; and
- Access to statistical, geographic and visual analysis tools.

Use of EMAP Information Management in Meeting These Requirements

EMAP IM will provide an integrated tool set for locating and accessing data throughout the EMAP program. Specifically EMAP IM will provide:

- A directory and catalog to index, summarize and provide detailed documentation of available data sets;
- Guidance on appropriate levels and types of metadata to collect during data collection activities by the EMAP Resource Groups; and
- Guidance on developing data quality indicators that are consistent across EMAP and other EPA research programs.

3. 2 Other EMAP IM Users

Other users outside the immediate EMAP program are considered in the design and development of EMAP IM. These users provide significant guidance, public support, and purpose for EMAP. While the scope of this Strategic Plan is limited primarily to Principal EMAP IM Users, other users of EMAP IM products will gain increasing importance in subsequent strategic planning activities.

Other users can receive data and information depending upon the nature of the specific requests. The Resource Groups and EMAP Assessment and Reporting functions are the main distributors of EMAP data to these other users, generally described as ***other scientists, management and policy makers,*** and the ***general public.*** Designated functions within EMAP IM will ensure that all EMAP and EPA policies are followed regarding data distribution.

3.2.1 Other Scientist Users

Background

There will be a large number of ***other scientists*** with backgrounds and needs similar to Principal EMAP IM Users. A variety of computing hardware and software will be used by these scientists, but they will typically use the Internet network for exchang-

ing messages with EMAP researchers, as well as locating and accessing EMAP data and information.

Requirements

These users will apply EMAP data and information in their own research and analysis. Primary IM requirements include:

- Ease in locating EMAP data and reports – because these users will not be experts in EMAP data, the ability to locate EMAP information will be crucial;
- Ability to identify metadata and other data documentation – needed in order to determine if EMAP data is applicable to their own research areas;
- Quality of data must be known and well documented;
- Data transfer standards must be in place to process information requests; and
- Access to tools.

Use of EMAP IM in Meeting These Requirements

EMAP IM will provide tools for locating and accessing data throughout the EMAP program. EMAP IM will:

- Develop and support a directory and catalog to index, summarize and provide detailed documentation of available data sets;
- Provide guidance on appropriate levels of metadata to collect during data collection activities by the EMAP Resource Groups;
- Provide guidance on developing data quality indicators that will be consistent across EMAP and other EPA research programs; and
- Provide guidance through the Interagency Data Interchange (IDI) function on methods for automating the initiation and servicing of information requests.

3.2.2 Management and Policy Making Users

Background

Policy makers and high-level managers may lack training in either environmental sciences or information technology. Typically, they will use personal computers for data analysis and presentation development. Requests for data often will be made by phone or other personal contact rather than through direct computer access.

Requirements

This user category primarily will be interested in accessing highly aggregated EMAP data. Primary IM requirements include:

- Easy access to high-level aggregated data and reports;

- Strong focus on certain geographic regions or particular time domains, which will require support for geographic information system (GIS) analysis; and
- Quality of data must be known.

Use of EMAP IM in Meeting These Requirements

Descriptions and locations of EMAP summary data will be stored in the EMAP Virtual Repository. EMAP Coordinating Groups can use this repository to handle external information requests. (For example, a "data warehouse," as described in Section 5, will provide users with consolidated data and information in an accessible format.) EMAP IM will develop and integrate GIS analytical tools that can display information for a specific spatial or temporal domain. Further, EMAP IM will provide guidance on developing data quality indicators that are consistent across EMAP. These indicators will be referenced in the EMAP Virtual Repository to determine data quality.

3.2.3 General Public Users

Background

EMAP information will become increasingly available to the *general public* within the five year period of this Plan. These users will have a wide spectrum of knowledge, training, and computer resources.

Requirements

The public will have a broad range of requirements, similar to the other categories of users, but at a much higher, more aggregated level. In particular, IM their requirements are expected to include:

- Browser and pointer interfaces to data and information;
- Easy access to high-level aggregated data and reports;
- Information bulletin boards;
- Strong focus on certain geographic regions; and
- Confidence that the information is correct and current.

Use of EMAP IM in Meeting These Requirements

Descriptions and locations of EMAP summary data will be stored in the EMAP Virtual Repository. Data and information sources such as the EPA Center for Environmental Statistics will be publicly available. If direct computer access is available, public users can access the Virtual Repository to locate and access data, metadata and information. EMAP IM will provide guidance on developing data quality indicators that are consistent across EMAP. These indicators will be refer-

enced in the EMAP Virtual Repository so that data and information quality can be determined.

3.3 Functional Requirements

A report on a National Science Foundation (NSF) workshop on scientific data management[13] stated: "There are some relatively simple questions that must be answered [from both computer science and scientific perspectives,] in order to enhance the scientific research environment":

- What data are available to me?
- What data have been collected?
- How can I get the data?
- How were the data collected/analyzed?
- Where are the data located?
- Where were the data collected?
- Who manipulated the data?
- Who collected the data?
- How good were collection methods?
- How good are these data?
- When can I get the data?
- When were the data collected?
- Why do I need the data?
- Why were the data collected?

EMAP users have indicated that these questions help define the basic set of user requirements. These questions lead to a set of guiding principles for EMAP IM systems development and the use of a formalized Architectural Framework, emphasizing the user's role. Figure 3.3.1 is one representation of functional requirements depicting providers of EMAP data on the left and users of processed data on the right.

Data Collection. This process acquires data from field samples, laboratory analysis, aerial or satellite images. Data collected over time usually will come in different formats, using varying collection equipment, staff, and analytical methods. In order to aggregate data collected from different regions or at different times, methods must be established and employed consistently to resolve differences to ensure meaningful summaries. These methods will change over the program's duration. Once the data have a common foundation (not necessarily common format or location), a directory and catalog of information can be established and distributed.

Data Verification. Collected data must be verified with appropriate data entry functions to ensure that they accurately reflect actual measurements, readings, observations, and analytical results.

Data Validation. This usually involves the comparison of related data over time. These data must be validated to ensure that the instrument or analysis is operating correctly. Scientists directly involved in obtaining the data should conduct the

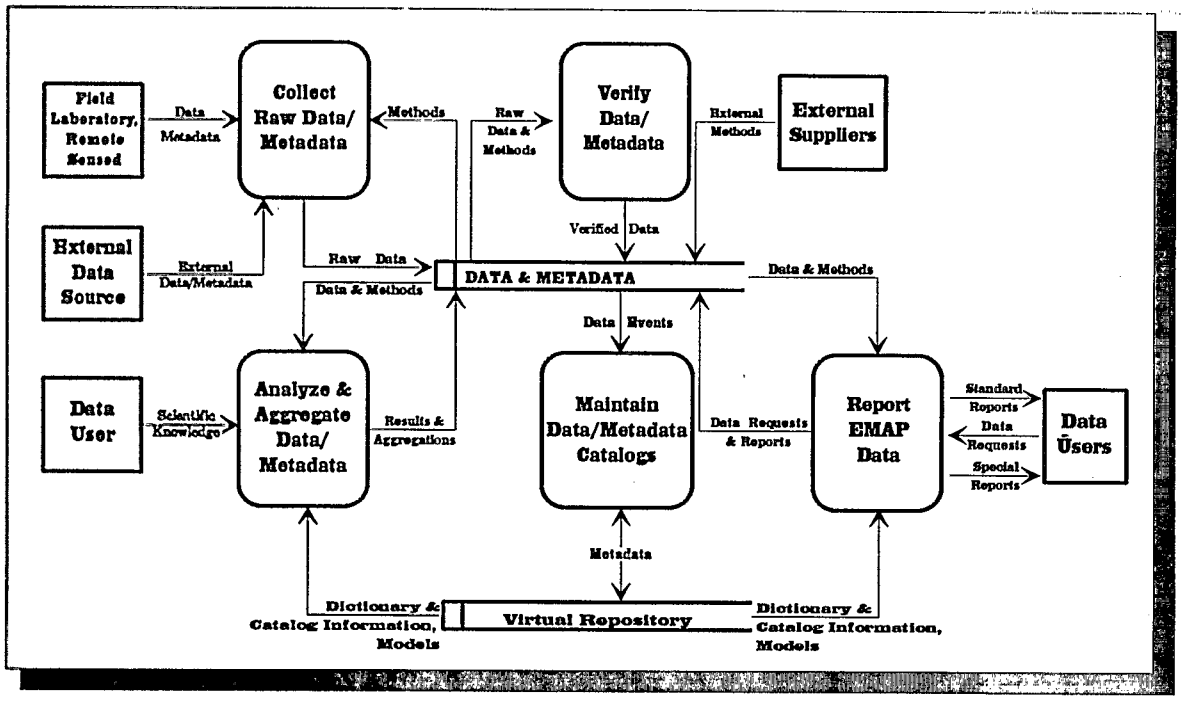

Figure 3.3.1. EMAP IM data flow diagram.

validation. Validation also includes assessing the method used to collect and process the data. Inappropriate methods could invalidate the data.

Data Aggregation. Certain data, when combined, serve to provide a new data element, for example, integrating dissolved oxygen, salinity, inorganic concentrations, and benthic abundance in order to produce a "benthic index". While seemingly a simple process, data aggregation actually requires ongoing planning, review, and validation. Strategies for aggregation range from creating and storing all possible aggregations to processing aggregations on request. Most often, a defined set of aggregations will be formed to produce a few routine reports such as annual reports. The capability to process aggregations on request, however, is invaluable in producing ad hoc reports.

Aggregations themselves must become part of the information resource within EMAP IM systems—a significant information challenge since the potential for millions of aggregations exists. Experience has shown that the effort in aggregation management, search, and retrieval is equal to or greater than the original data collection. Data aggregation may involve combining EMAP data with supporting data. In order for this process to be useful, it must be documented so it can be repeated accurately. Figure 3.3.2 illustrates several levels of data aggregation showing how it matures from simple data points to result in an information product, (e.g., a final report).

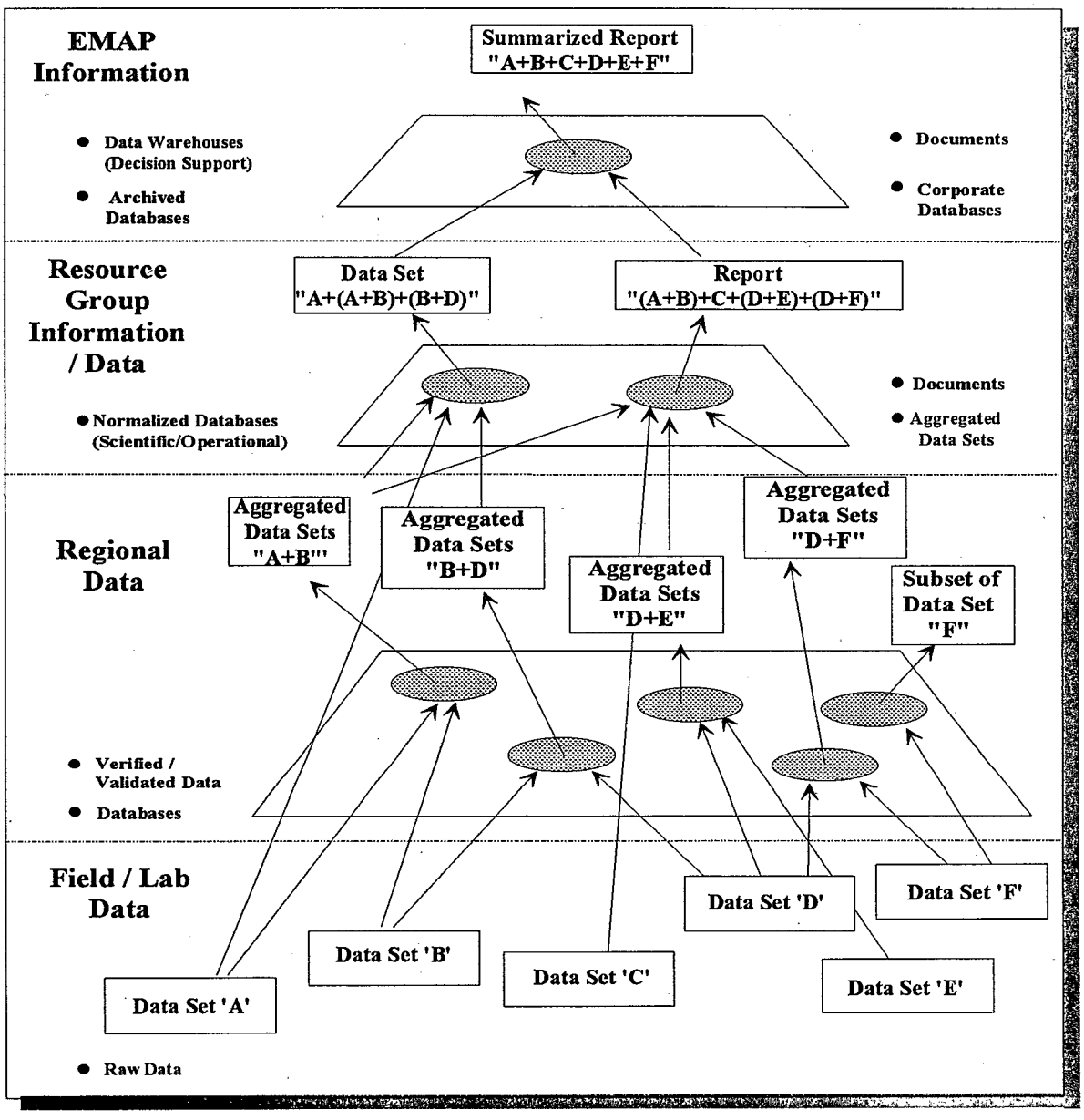

Figure 3.3.2. Multi-layer data/information aggregation.

Data Integration. The integration of response, exposure, and stressor data at the resource group level for the compilation of the annual statistical summary is currently the primary goal within EMAP. Integration across resource groups is equally challenging, and in some cases more ecologically interesting, and was one of the original expectations for the program. Successful integration across resource groups will be one test of the EMAP IM system.

Data Distribution. This occurs throughout the data maturation process; however, it is not possible to acquire the data without documenting the descriptive support information. Data distribution will be in paper and digital forms.

Data Archive. This process must be established to ensure no loss of data during the entire process of maturation. A fundamental rule should be employed: *It is easier to recover from bad analysis than from data of unknown quality or lost data.*

Data Storage. Strategic planning for the development and implementation of IM systems requires an estimate of the volume of data, as shown in Table 3.3.1. To derive these estimates, data storage requirements were extrapolated using two summer pilot data collection activities from the Estuaries Resource Group.[14] The numbers are in gigabytes and include data, descriptions of the data, and initial aggregations for individual ecological resources, and include spatial data for the Resource Groups. This analysis predicts that requirements for database capacity will increase approximately 10-fold in 5 years.

FY 93	FY 94	FY 95	FY 96	FY 97
16.7	30	55	90	150

Table 3.3.1. Estimated EMAP IM Database Size in Gigabytes

However, quantity is not the greatest data storage challenge for EMAP IM. Diversity of data will have a greater impact on IM resources than volume of data. EMAP IM must deal with complexity of data and metadata, the relationship of data and metadata, the derived aggregation of data and associated metadata, spatial data, and new technology in deriving data, and other factors. these variables will offer IM challenges that exceed any challenge posed solely by present or projected data volume.

3.4 System Requirements

The EMAP IM systems are scientific information systems. These systems must support significant descriptive information about data for yet unknown uses, and must allow for interfaces to new analytical methods. They must accommodate integration of related information and provide the scientist timely access to data that is widely distributed among elements of the information system. In addition, IM systems must allow for ongoing change in the information, engineering, and computing sciences.[15] In general the information system should include but not be limited to the following:

- Data acquisition, documentation, and storage;
- Distributed but coordinated information management allowing access to integrated data and information transfer in multiple formats;
- Maintenance of data integrity and security;

- Integrated analysis tool and methods;
- Flexibility and ease of use in accessing and exchanging data and information;
- Efficient interoperability between Resource and Coordinating Groups; and
- Flexible, powerful information reporting and display tools.

3.5 Additional Considerations

At present, certain constraints limit the scope of what can be accomplished by EMAP IM. These constraints apply to process and technology and are represented below.

Distribution of EMAP Data Outside of EMAP

Distribution of EMAP data outside of EMAP will depend upon available hardware, software, and communications links. Integration requirements for hardware, software, and communications must still be identified. This challenge will be addressed before integration with outside data begins.

Access and Inclusion of Data From Outside Sources

Inclusion of data from outside sources depends upon available hardware, software, and communications links. Additionally, since the data is owned and maintained by an external source, the stability of the format and contents of these data is not under EMAP control. Extra effort may be required in order to access data of different formats or platforms. Requirements and challenges associated with this constraint must be resolved before external data can be integrated into EMAP.

Mass Storage Systems and Long-term Archival

A planning factor for EMAP data is that it must be kept for at least 20 years. Therefore, a reliable storage and retrieval system for large amounts of data is critical. Most Resource Groups probably will not have the capabilities for long-term storage of large amounts of data; therefore, EMAP Central or EPA facilities will be used. EPA facilities contain procedures and capabilities to deal with this constraint. Therefore, it is expected that Resource Group data will be transferred to such central archive facilities. CD-ROM technology is expected to provide this long-term storage and will also provide fairly fast access. Identification of data to be archived still must be determined as EMAP approaches the Enterprise Implementation process.

Various Types of Database and File Systems

In order to be useful to a wide variety of users, EMAP must be able to communicate with a wide variety of computer types and software packages. Presently, there are anticipated problems with communications among different types of file systems

and software and problems due to incompatibilities between different versions of software. The current database management system for EMAP IM is Oracle. Oracle uses a structured query language (SQL) allowing it to interface with other relational database systems (RDBMS) that also use the SQL standard. Other types of non RDBMS systems are also in use throughout EMAP, such as SAS and ASCII files. Integration between different file systems and data formats is being addressed in the Proof-of-Concept which is explained in Section 5.

Security

EMAP must protect the confidentiality of some data, and must be able to protect all data from damage due to: external and environmental threats; hardware and software error; operations or procedural error; and, malicious actions. Procedures to address this issue are being considered by the QA/QC Coordinating Group.

System Dictionary, Directory and Catalog

A system dictionary, directory and catalog are critical for keeping track of the large amounts of data that EMAP will contain. Presently, there are issues regarding how to keep distributed files in synchronization, how to allow for easy update while maintaining integrity, how to keep in synchronization with changing data, and how to keep metadata synchronized with data sets. These issues will be addressed by EMAP IM in the design, development and implementation of the Virtual Repository.

Advanced User Interface

User interface technology is evolving at a rapid rate. The EMAP IM architecture is designed to be sufficiently flexible to allow use of current technology yet be capable of adapting to future technology. This will be an on-going issue with EMAP IM.

Processing Distribution

Processing resources must be distributed with regard to current hardware and software technology. Consideration must be given to network availability and capacity, server speed and capacity, and availability and functionality of user workstations. This issue is being addressed in the Proof-of-Concept.

Support for a System That Will Last Decades

Data that has been stored must be retrievable and meaningful. It is important that the metadata describing the data parallel the quality of that data. Also, audit trails to record procedures that were performed on data are important. These issues are currently being addressed by EMAP IM.

Section 4

GUIDING PRINCIPLES FOR EMAP IM

Large information systems that evolve over time are inherently complex. This can result in system developers losing focus on initial principles considered fundamental for continued project success. The guiding principles contained in this Section are statements for directing EMAP information management (IM) systems development. These principles adhere to National data management policy endorsed by Dr. Allan Bromley as Director, Office of Science and Technology Policy[16]and the current National Information Infrastructure initiative.[2] These guidelines were selected because they pertain to the scientific focus and functionality of EMAP. They will affect all facets of EMAP IM systems development including ***policy, data, design, operations, technology,*** and ***users***. They also provide guidelines for project management and serve as a basis for the technical approach. These guiding principles will influence every EMAP manager, developer, and user of EMAP IM systems and provide appropriate focus and boundaries.

4.1 Guiding Principles for Information Management

The quality of information must be known and reasonable.

EMAP IM systems must track the integrity, quality, and pedigree of all EMAP data throughout its life cycle. The data must be of known quality before its inclusion into EMAP IM systems. [17-23] As the data is verified and analyzed, the information system will track data quality. Analysis often involves data aggregation, which may require the formulation of new methods for determining aggregated data quality. Since data collection is distributed and occurs in different ecological resources, the process is highly complex and will require substantial coordination and cooperation among EMAP Resource Groups to ensure knowledge of data location and quality at all times.

Management of large complex data requires automated systems.

EMAP IM systems must manage and store data efficiently. Because EMAP is a National program, its IM system must plan for automated monitoring through large, widely-distributed data collection activities that, wherever possible, must adhere to common data standards. Also, EMAP IM systems must provide automated information tracking.[24-27] Data processing activities that include verification, validation, and analysis should also be automated whenever possible. Routine analysis and report generation should be automated, but should not impede user access and verification of data quality.

Change is inevitable and continuous.

EMAP IM systems must plan for and manage changes in scope, objectives, user requirements, technologies, data, and personnel. Correspondingly, EMAP IM must be largely independent of specific technologies because requirements and technology are constantly evolving.

Science and EPA policy will determine access to EMAP data.

Access to data will be in accordance with EPA policy. EMAP IM systems must provide controlled access to EMAP data because of the diversity of intended users. EMAP IM systems must support Environmental Protection Agency (EPA) entities and other Federal agencies in attaining appropriate access. The level of access will be evaluated by the organization responsible for the data in order to appropriately accommodate the needs of requesting parties.

4.2 Guiding Principles for Policy

EMAP IM must be a catalyst for information standards.

For EMAP IM systems to succeed, they must be built upon information standards. EMAP IM will actively identify, prototype, and support information standards within EMAP, EPA, and other agencies and organizations that produce such standards.

EMAP will adopt preexisting IM policies where possible.

EMAP IM will establish procedures for implementing policies related to data confidentiality, data access, and data distribution. EMAP IM will examine existing EPA, governmental, and scientific information management policies, incorporating existing procedures that fulfill EMAP IM system and user needs. If necessary, EMAP will formulate its own IM policies to ensure that EMAP-specific IM needs are met.

4.3 Guiding Principles for Data

EMAP information is comprised of data and metadata.

Data are logically and functionally inseparable from metadata. While raw data has little value without descriptions (e.g., sampling location, sampling technique, analysis methods, or quality control procedures), metadata are vital information and can be useful when analyzed or reported even without associated data. For example, a summary report of 100 samples without knowing the sampling technique may not be valid, but a report on the sampling technique could provide useful information. Data must not be distributed or reported without associated metadata also being available. Like the data, metadata are diverse, distributed, and changeable over time.

Information integrity must be managed.

Data and metadata integrity must be maintained. This includes a common definition and tracking of information pedigree, archival, quality, and physical security (including backup and recovery).

Information integration and distribution requires standards.

Data conventions and standards must be chosen and followed. The development of these standards are the responsibility of scientists, engineers, and management. The standards should include such topics as metadata codification, information exchange protocols, analysis methods, and reporting methods. While local analysis can be performed with meaningful results, integrated and decentralized data sources that may change with time require standards.

Analytical access to data must be provided.

The value of EMAP data relates directly to the types of data analysis tools that are available and their ease of use. EMAP IM systems will provide an integrated environment that includes interfaces between data storage and suites of analytical programs. As data are processed, the analytical history will be captured in metadata files which allows for the regeneration of computational results.

4.4 Guiding Principles for Design

IM systems must match the scientific objectives of EMAP.

EMAP IM must propose, develop, and maintain the most appropriate information systems architecture that will best meet EMAP's scientific objectives. This must be accomplished within the larger context of current EPA standards and policies. EMAP IM must develop and evaluate prototypes and alternative approaches when current standards are not appropriate.

IM systems must accommodate growing and changing information.

EMAP IM systems must have a flexible architecture to accommodate the variety and large volume of data continually being collected. Systems must not only store the data, but also must track changes in the data.

IM systems must be formal, yet flexible in design and implementation.

EMAP IM systems must be reliable, maintainable, secure, and well-documented. Due to the need for ongoing change, IM systems must strike a balance between formal design methods that typically require substantial time and documentation, and prototypes that use Rapid Application Development (RAD) techniques.

IM systems must facilitate inter-agency cooperation.

EMAP IM must pay explicit attention to related and external information systems to allow sharing and interoperability where possible. Ancillary data must be evaluated for relevance to EMAP to preclude possible redundancy or omission of applicable data.

IM systems requirements will be described in a technology-neutral fashion.

Information systems requirements can be difficult and expensive to gather. In order to accommodate the continuous change in technology, user requirements will be gathered, documented, and maintained in technology-neutral terms.

4.5 Guiding Principles for Operations

Operations must provide secure, quality service and regular review.

EMAP IM must ensure that Operations (the agent or agents which provide customer service and support) is responsive to both user and systems needs. Operations must continually monitor and improve the quality of service in such areas as documentation, training, user support, and systems review.

Operations must be responsive to users.

EMAP IM must ensure that Operations accommodates the needs of EMAP IM users. Initially, these users will be the Resource and Coordinating Groups, Cooperative Partners, and other principal users. Operations should be demand-driven and able to respond to unscheduled requests for information when appropriate. Operations must provide feedback to systems development to ensure continuous process improvement.

Operations must meet product needs.

EMAP IM Operations must ensure that appropriate Resource and Coordinating Group users have the IM resources necessary to produce information products, including reports, assessments, and presentations, consistent with EMAP priorities and data access policies.

Operations must facilitate data sharing.

EMAP IM must ensure that Operations facilitates data access, sharing, integration, and interoperability among designated EMAP groups and external users.

Operations must enable effective communication of project and scientific information.

EMAP IM must ensure that Operations facilitates project and scientific communication to fulfill objectives and to communicate project activities, meetings, staff directories, reports, publications, and collected scientific data.

4.6 Guiding Principles for Technology

EMAP will use the best practical technologies.

EMAP IM systems will integrate many separate software and hardware components into a comprehensive infrastructure. Some components will be developed and customized specifically for the EMAP project, but most will be available from commercial sources. EMAP will work with industry, universities, and the National laboratories to develop new technologies that improve EMAP IM systems. Where applicable and practical, new technologies will be evaluated and incorporated.

Technology will constantly change.

EMAP IM must accommodate the changes which will occur in information systems, scientific analysis components, communications, and other relevant aspects of the EMAP program. But EMAP will not incorporate these changes until they are deemed stable and suitable for use.

4.7 Guiding Principles for Users

The EMAP user community is large and has diverse needs.

EMAP IM systems must be designed for a wide range of users while also allowing specific, special purpose use. Systems must be highly flexible and adaptable to individual needs, and they must run on a variety of geographically dispersed and diverse hardware platforms.

EMAP users must be provided timely access to EMAP data.

EMAP IM systems must facilitate timely access for all approved users, especially the initial users (Resource Groups), whose responsibility is to collect, validate, and summarize the data before distribution.

Section 5

TECHNICAL APPROACH

5.1 Introduction

The technical approach for developing Environmental Monitoring and Assessment Program (EMAP) information management (IM) systems consists of two main elements:

- A systems engineering approach for EMAP IM systems design and development, and
- The data, process, technology, and network architectures that describe the key components of EMAP IM.

The systems engineering approach for EMAP IM is an evolutionary approach using elements of other "formal" systems development approaches.[28-42] This evolutionary approach uses the Zachman Framework, illustrated in Table 5.1.1, as the architectural "checklist" or guide to ensure that the full breadth of requirements are considered, in an *Enterprise* context, during the systems life cycle.[3] Joint application design (JAD)[43] and rapid application development (RAD)[6] methods will be used in the EMAP IM hybrid evolutionary approach.

5.2 Systems Engineering Approach

Scientific information systems development is an evolving engineering process that requires specific considerations in order to succeed. A systems development methodology or approach is a principal consideration. Different approaches that historically have been used for systems development include: top-down

	Data	Function	Network	People	Time	Motivation
Scope	List of things important to the "business"	List of processes the "business" performs	List of locations in which the "business" operates	List of organizations or "agents" important to the "business"	List of events significant to the "business"	List of "business goals and strategies
Enterprise Model	Entity relationship diagram with business entities and business constraints	Process flow diagram with business processes and business resources	Logistics network	Organization chart showing organization units and work product	Master schedule of business event and business cycles	Business Plan with objectives and strategies
Information System Model	Data model with data entity and data relationships	Data flow diagram showing application function and user views	Distributed system architecture	Presentation architecture with roles and deliverables	Processing structure	Knowledge architecture
Technology Model	Data design	Structure chart with computer function and screen formats	System architecture with hardware/ software and line specifications	User interface architecture	Control structure	Knowledge design
Components	Data definitions/ descriptions	Software program	Network architecture showing node addresses and protocols	Security architecture	Timing definition	Knowledge definition
Functioning System	Data	Function	Network	Organization	Schedule	Strategy

Table 5.1.1. Zachman Framework

approaches, bottom-up approaches, transactions-based approaches, knowledge engineering approaches, rapid-prototyping approaches, and evolutionary development approaches. While most approaches or combinations of these approaches can succeed, many have failed because of a lack of interaction among the people working on different parts of the system. For EMAP IM, a hybrid evolutionary approach has been selected using components of several formal system development approaches.

5.2.1 A Formal System Development Approach

The system development life cycle (SDLC) consists of five major phases.[5] A brief summary of each life cycle phase follows:

- **Initiation** – Identifies the information management problem to be solved with a focus on the pertinent information, organizations experiencing the problem, time frame available for establishing the solution, and overall value of the solution.
- **Concept** – Provides a high-level, comprehensive model of the solution to the problem that will guide the effort in subsequent phases. This phase defines

high-level functional and data requirements, and evaluates alternative solutions to these requirements. The solutions address all aspects of the system: the information to be processed; functional processing capabilities; hardware, software, and communications to be used; and project organization and staffing through the end of the system life cycle.

- **Definition and Design** – Provides a detailed description of the information and processing capabilities required of the system, and subsequently a detailed description of how the system will provide these capabilities. This phase addresses the details of manual procedures as well as automated components of the system.
- **Development and Implementation** – Acquires or builds the system in accordance with the prescribed design and installs the system in the production environment in which it will be available to the users. This phase also incorporates needed data into the new system and trains users and system support staff prior to the start of full system operation.
- **Operations** – Provides the full capabilities of the system to the users and ensures adequate ongoing maintenance support for the system. This phase includes system modifications, periodic formal evaluations of the system, feed-back, and the ultimate termination and archival of the system at the end of its useful life.

5.2.2 EMAP IM Evolutionary Approach

Merely completing each phase of the SDLC process has not proven totally successful in ensuring that operational systems are fully responsive to user requirements. One primary limitation has been the lack of user involvement and feedback during the development and testing phases prior to implementation. In other words, initial user needs are not met in the production system. To avoid this and other shortcomings of earlier more traditional systems engineering approaches, EMAP IM has adopted a hybrid evolutionary approach to systems development which adapts to evolving user needs and an evolving technology base.

One of the better known alternative system development approaches that support evolutionary development is Barry Boehm's Spiral Model,[44] as modified for EMAP IM and shown in Figure 5.2.2.1. The Spiral Model supports the concept of "rapid prototyping," that is, building smaller, less complex versions of the system to demonstrate specific aspects of the system. Demonstration work could focus, for example, on the human-machine interface, the integration of different software packages into a comprehensive solution to a particular problem, or the feasibility of using a particular technology as part of the system.[45]

This approach allows for ongoing use of operational systems while the development process occurs. For EMAP IM, the design process occurs in rapid series with an evolutionary development process that will reuse previously developed system components. These components are expected to have real uses and stimulate direct and

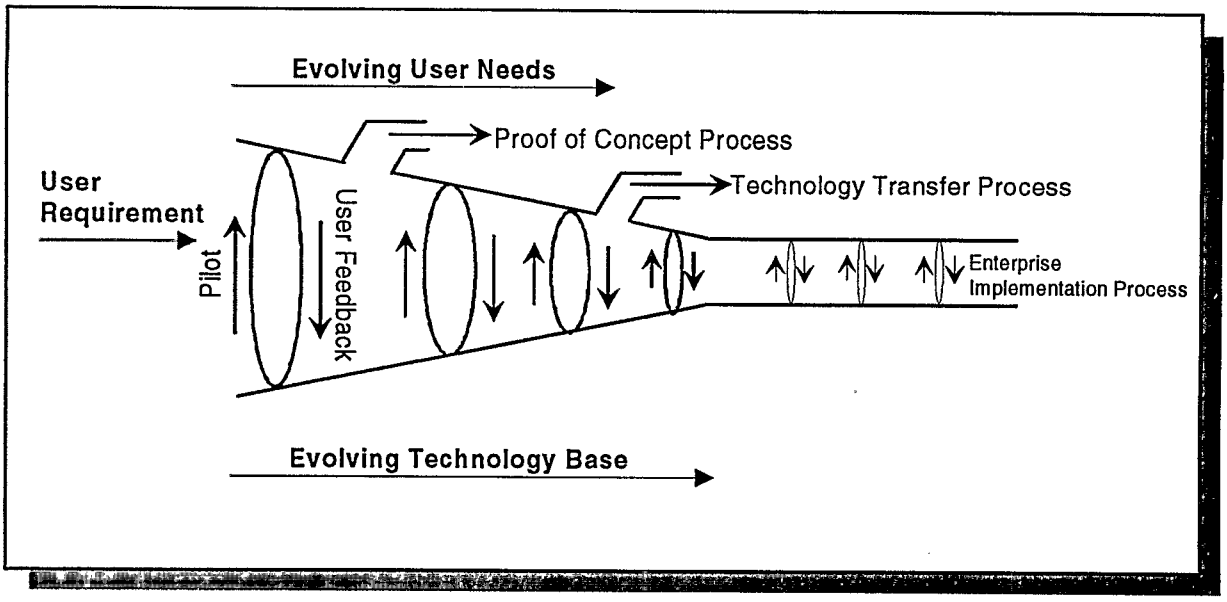

Figure 5.2.2.1. EMAP IM evolutionary process.

early user feedback through "Proof-of-Concept" versions. These will be sufficiently developed so that users can apply the component with little direct involvement of the implementation team. For example, a "Proof-of-Concept" user interface is being developed early in the life cycle. EMAP users will be able to employ this interface to perform limited real-world functions. Based on user feedback, various iterations of the user interface will be produced with successively greater functionality. The technology of this improved interface would be transferred to other users who, in turn, continue to provide feedback.

5.2.3 Architecture Framework

The EMAP IM Architecture Framework uses a modified version of the Zachman Framework illustrated in Table 5.1.1 to assist in the system development process. The following explanation of the Framework is taken from *Extending and Formalizing the Framework for Information Systems Architecture.*[46]

> [An Enterprise] contains entities, processes, locations, people, times, and purposes. Computer systems are filled with bits, bytes, numbers, and the programs that manipulate them. If the computer is to do anything useful, the concrete things in the [Enterprise] must be related to the abstract bits in the computer. [A] framework for information systems architecture (ISA) makes that link. It provides a systematic taxonomy of concepts for relating [elements of] the Enterprise to the representations in the computer. It is not a replacement for other programming tools, techniques, or methodologies. Instead, it provides a way of viewing a system from many different perspectives and showing how they are all related.
>
> The purpose of the [Architecture] Framework is to show how everything fits together. It is a taxonomy with 30 boxes or cells organized into six columns and

> five rows. (The sixth row is the result of the five above it.) . . . Flow charts, for example, may be suitable for describing the cell in the function column, components row in the Framework; and entity-relationship diagrams (ERD) may be acceptable for the data column, system model row. But the Framework shows how the cells in different columns and rows relate to one another.
>
> When applied to an information system, the word architecture is a metaphor that compares the construction of a computer system to the construction of a building. The Zachman Framework is an elaboration of the metaphor. It compares the perspectives in describing an information system to the perspectives produced by an architect in designing and constructing a building.

The Framework shows that the information system development process requires several levels of architectures. As depicted in Table 5.1.1, the six Framework architectures are: ***Data, Function, Network, People, Time,*** and ***Motivation.*** The levels of the Framework are: ***Scope, Enterprise Model, Information System Model, Technology Model,*** and ***Components.*** Each Framework architecture (e.g., Data) develops as the process progresses though the Framework's life cycle from ***Scope*** through ***Components.*** This process development cycle is applied to each architecture within the Framework as depicted in Figure 5.2.3.1. This iterative process will be applied to each subsequent Enterprise component during the three IM implementation processes: ***Proof-of-Concept (POC), Technology Transfer,*** and ***Enterprise Implementation.*** Following is a description of each framework level, using a building construction project as an example. The first three levels provide a technology-independent model to accommodate emerging hardware development and software changes.

- **Scope.** The first architectural sketch would depict in gross terms the size, shape, spatial relationships, and basic purpose of the final structure. In the Framework, it corresponds to an executive summary for the planners of the EMAP System: what it will cost, and how it will perform.
- **Enterprise or Business Model.** Next are the architect's drawings that depict the final building from the perspective of the owner, who will have to live with it in the daily routines of business. They correspond to the Enterprise model, which constitutes the design of the EMAP system and shows the basic entities and process and how they interact.
- **System Model.** The architect's plans are the translation of the drawings into detailed specifications from the designer's perspective. They correspond to the system model designed by systems analysts who must determine the data elements and functions that represent EMAP entities and processes.
- **Technology Model.** The contractor must redraw the architect's plans to represent the builder's perspective, which must consider the constraints of tools, technology, and materials. The builder's plans correspond to the technology

model, which must adapt the EMAP information system model to the details of the programming languages, databases, communications, or other technology.

- **Components.** Subcontractors work from shop plans that specify the details of parts or subsections. These correspond to the detailed specifications that are given to Information Managers and programmers who develop individual database systems and other components without being concerned with the overall context or structure of the system rows of the Framework.

Using the Framework checklist ensures that all aspects of the system are addressed from a total system or "Enterprise Approach" and that all participants in the development and implementation of the Enterprise have a common reference. (This Framework also has been adopted for use by the EPA Office of Information Resources Management [OIRM] as the basis for an EPA-wide architectural approach.)

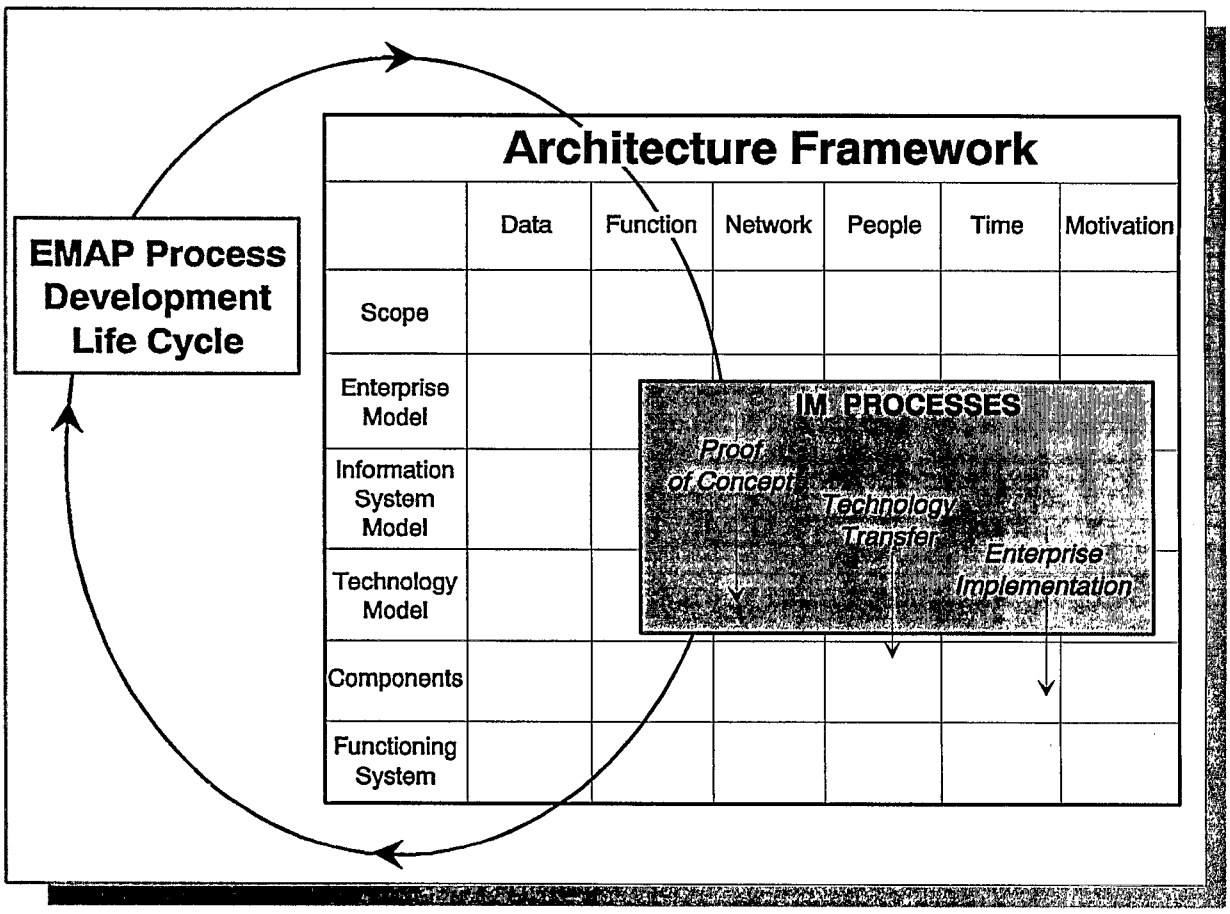

Figure 5.2.3.1. Architecture framework and the evolutionary process.

5.3 System Architecture

The EMAP IM System Architecture provides the basis for developing information management systems that can accommodate the rapid and continuous changes in technology. As discussed in Section 5.2.3, the six Framework architectures are:

- Data architecture;
- Function architecture;
- Network architecture;
- People architecture;
- Time architecture; and
- Motivation architecture.

5.3.1 Data Architecture

To fully determine EMAP data requirements, an Enterprise-wide data architecture will be produced using the Architecture Framework. This architecture will be used to ensure that data and associated metadata from the various EMAP data sources will be stored, managed, and retrieved in a consistent manner with known quality. This data architecture will also be used to facilitate information exchange within EMAP, EPA, and other programs and scientists. The structure of the data architecture is in terms of "entity-relationship-entity," and may be thought of as consisting of two layers: the *Data Layer*, and the *Virtual Repository Layer*.

Data Layer

Data at the Resource Group nodes consists of scientific data (field data sets and laboratory analysis data sets), metadata and documentation describing the scientific data, and external data sets used for analysis of the data. These data sets are stored in electronic form such as SAS data sets, ASCII files, and various databases.

At each Resource Group node, some of the data sets will be put into a relational database system (RDBMS). These databases may also be referred to as *Operational Databases*. Determination of which data sets are put into the operational databases will be made by Resource Group and EMAP Central personnel. The database tables will be in third normal form, i.e., no repeating information in each table.

EMAP Central will store EMAP data sets, external data sets, and documents of interest to the entire EMAP project. Databases at the Central node will be organized into three categories: 1) Warehouse, 2) Corporate, and 3) Archived.

The *data warehouse* is a database used for decision support that is physically separated from the operational and corporate databases. A data warehouse is created by extracting information from multiple operational databases and consolidating the

information into a more accessible format. Using a data warehouse, with its consolidated data, minimizes system performance degradation caused by ad hoc queries and resource-intensive data summaries and aggregations.[47] The data warehouse can also be used to store information products that are produced during these data analysis activities.

The *corporate* databases will serve the Coordinating Groups. These databases contain information to be shared system wide and will be stored at EMAP Central.

The *archived* databases refer to copies of the operational databases owned and maintained by the Resource Groups. These archived databases may be stored at EMAP Central in order to support disaster recovery for Resource Groups that do not possess adequate disaster recovery facilities.

Virtual Repository Layer

The Virtual Repository is a collection of metadata describing EMAP data and information systems, plus a set of tools that makes metadata available to users and system developers. The Virtual Repository will be distributed throughout the EMAP IM system. Each node will contain a part of the Virtual Repository that will point to and access data and information on that node, and other nodes. The parts of the Virtual Repository are described below.

- The **Model Manager** is a tool that manages models, such as entity-relationship diagrams, data flow diagrams, function hierarchies, and database schemas. These are objects used to design and understand a database system. The model manager will store these objects, keeping track of different versions, provide access to the objects with check-in/check-out synchronization, convert from one format to another, and assist in model integration.
- The **Data Dictionary** contains descriptions, formats, and other basic information about items in the EMAP database system.
- The **Configuration Manager** is a tool that assists in keeping track of versions of models, software, etc.
- **CASE Tools** provide graphical and textual interfaces to the objects contained in the model manager and data dictionary.
- The **Directory** provides the means to index and track data sets, and provides summaries of these data sets.
- The **Catalog** locates and provides access to detailed documentation on the data sets.

Figure 5.3.1.1 illustrates the components of the data layer and the virtual repository layer, and the relationships between the virtual repository and data layers.

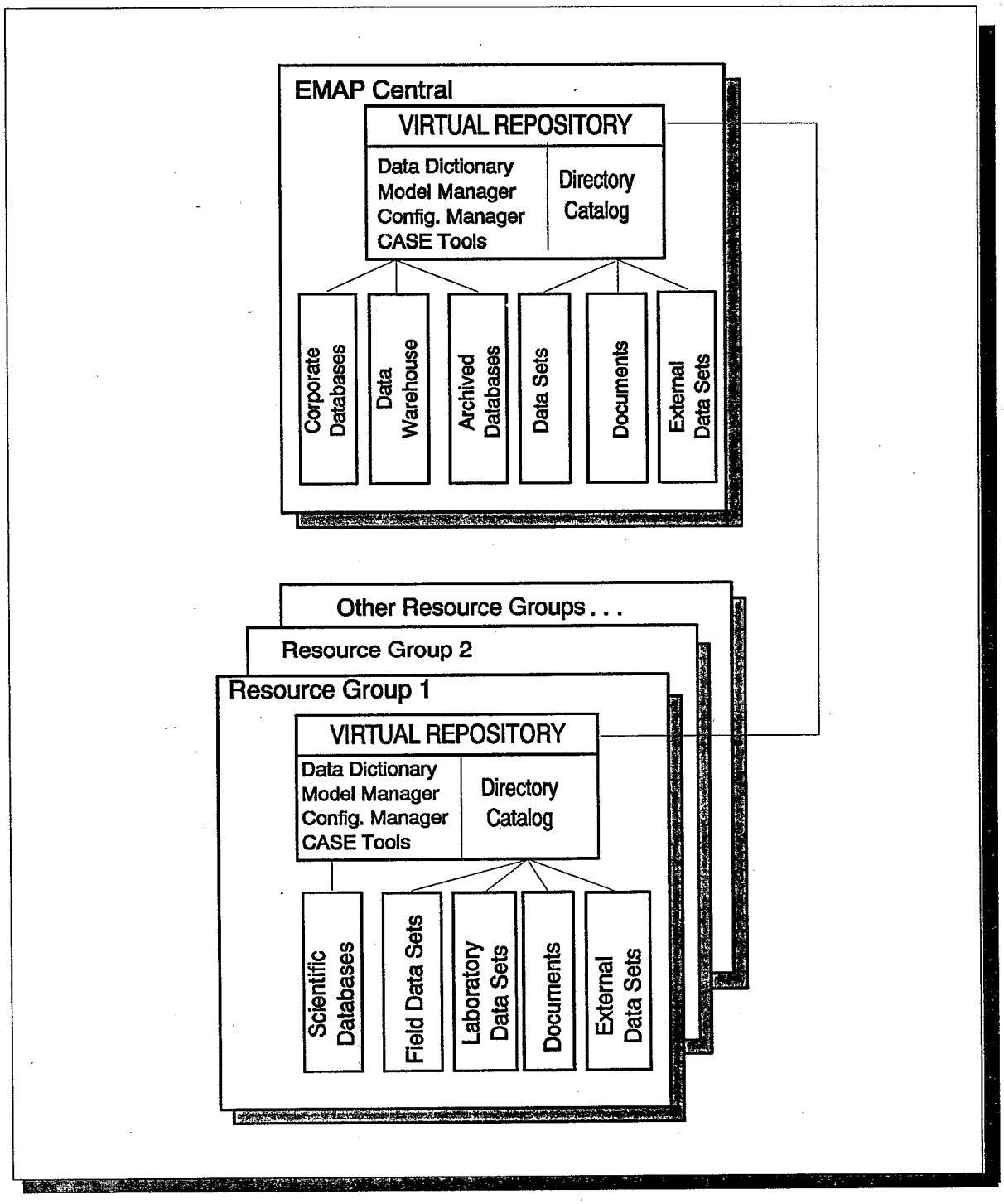

Figure 5.3.1.1. EMAP virtual repository architecture.

Data Sources

Data in EMAP will come from a variety of sources and be used for different analytical purposes as Figure 5.3.1.2 illustrates. Based on EMAP IM planning meetings, different data are identified that will be generated and used by EMAP. These data are described in the following text. Table 5.3.1.1 provides examples of some of the metadata associated with each type of data mentioned below.

- **Raw monitoring data** – measurements and observations taken in the field.
- **Verified monitoring data** – raw monitoring data that has been verified. Verification refers to the process of determining whether procedures, processes, data, or documentation conform to specified requirements. Verification activities may include inspections, audits, surveillance, or technical review.
- **Validated monitoring data** – monitoring data that has been validated. Validation refers to the systematic process of reviewing a body of data against a set of criteria to provide assurance that the data are acceptable for the intended use.
- **Aggregated monitoring data** – monitoring data that has been summarized or transformed.
- **Laboratory methods** – protocols used in an analytical laboratory for determining a sample's constituents.

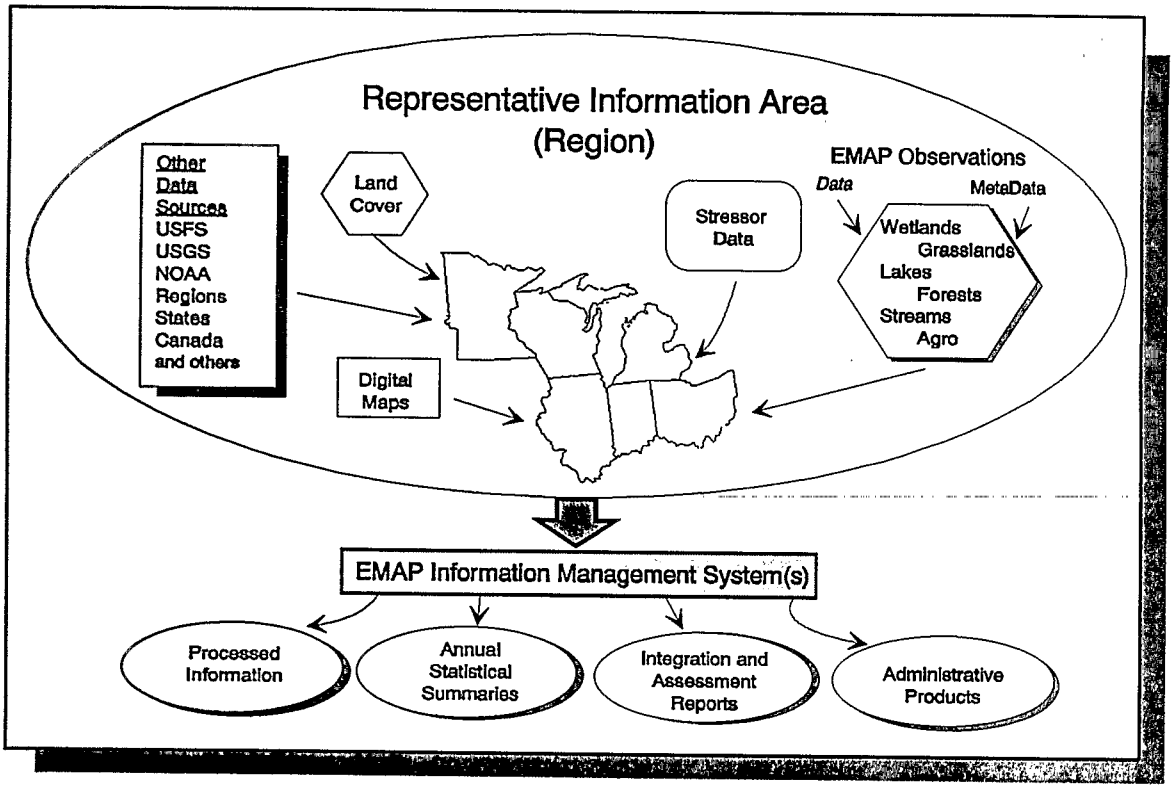

Figure 5.3.1.2. EMAP data flow.

Data Class	Examples of Scientific Metadata	Examples of Data Processing Metadata
Raw monitoring data	Field team name, date sample collected, weather conditions during sampling period, lab notebook number, data quality objectives. Located in Resource Group Field Data Set.	Name of data set created by automatic data logger. Format of data set. Located in Virtual Repository Directory.
Verified monitoring data	Verification methods used. Located in Resource Group Data Set.	Data set name, format of data, indicator whether data loaded into database system, database table name, and attributes of ERDs describing schemata, etc. Located in Virtual Repository Directory.
Validated monitoring data	Statistical methods used to detect outliers. Located in Resource Group Data Set.	Standard units of measure, numeric precision. Located in Virtual Repository Data Dictionary.
Aggregated monitoring data	Aggregation method. Located in Resource Group Data Set.	Name of aggregated data set or database table. Located in Virtual Repository Directory.
Laboratory methods	Analysis protocol, detection limits. Located in Directory and Catalog.	ERD of Laboratory Methods database. Located in CASE Tools in Virtual Repository.
Field methods	Sample collection and preparation instructions, calibration dates of portable monitoring equipment used. Located in Resource Group Data Set.	Indicator if methods conform to standards. Located in Virtual Repository Directory.
Geographic resource data	Resolution of remote sensing devices, position accuracy standards. Located in Resource Group Data Set.	Indicator if geographic data conforms to spatial metadata exchange standard. Located in Virtual Repository Directory.
Indicator methods	Aggregation methods. Located in Virtual Repository Directory and Catalog.	SQL statements for computing indicator (or Algorithms). Located in Resource Group Data Set.
Indicator measurements	Sampling period. Located in Resource Group Data Set.	Pictorial View of timing. Located in Virtual Repository Directory and Catalog.
Ancillary data	Taxonomic classification system. Located in External Data Set or Scientific Database at a Resource Group.	Version of the data set and when it was created. Located in Virtual Repository Directory and Catalog
Status and tracking data	Location of a data set while in the analysis process. Located in Virtual Repository Directory.	Indication of completeness of tracking data. Located in Virtual Repository Directory and Catalog
Office information	Date information reviewed / revised. Located in Virtual Repository Directory.	Procedures for receiving information. Located in Virtual Repository Directory and Catalog.
Reference documents	Document peer review date, accepted for publication flag. Located in Resource Group Documents "Folder".	Format of document. Located in Virtual Repository Directory.

Table 5.3.1.1. Examples of Metadata

- **Field methods** – protocols used in the field for taking observations or determining a sample's constituents.
- **Geographic resource data** – spatially-related information.
- **Indicator methods** – methods for constructing a metric that can measure the health of an ecosystem.
- **Indicator measurements** – calculated values of a specific ecosystem indicator.
- **Ancillary data** – miscellaneous data required for EMAP. This could include taxonomic data, chemical constituent information, etc.
- **Status and tracking data** – information related to the management of EMAP activities.
- **Office information** – administrative information.
- **Reference documents** – papers, articles, or other documentation.

5.3.2 Function Architecture

Along with the data architecture discussed in Section 5.3.1, an Enterprise-wide function architecture will be produced. Also, this function architecture will use the Framework mentioned in Section 5.2.3 as a checklist to ensure that all relevant EMAP IM functions are considered during the system life cycle. The function architecture will be used to ensure that the data architecture takes into account the data or metadata that is needed to perform the designated functions. Also, where appropriate, the function architecture can be used to verify that procedures used by the different EMAP groups are consistent.

Figure 5.3.2.1 illustrates the major interrelationships between these activities and the major classes of data mentioned in Section 5.3.1. As can be seen from the figure, the function architecture consists of "input-process-output" structures. For example, field and laboratory methods are input to the routine monitoring process, which generates raw monitoring data as output.

Some of the major types of functions that EMAP IM will support are:

- **Research done prior to monitoring** – scientific research to determine ecosystem indicators and monitoring approaches;
- **Routine monitoring** – regularly scheduled monitoring activities;
- **Assessment** – analysis and summarization of monitoring results;
- **Reporting** – publishing EMAP reports and articles;
- **Acquisition of external data** – importing data into EMAP from other sources;
- **Data distribution** – exporting EMAP data to other users;

- **Marketing/Training** – publicizing EMAP activities; and
- **Administrative** – management of EMAP activities.

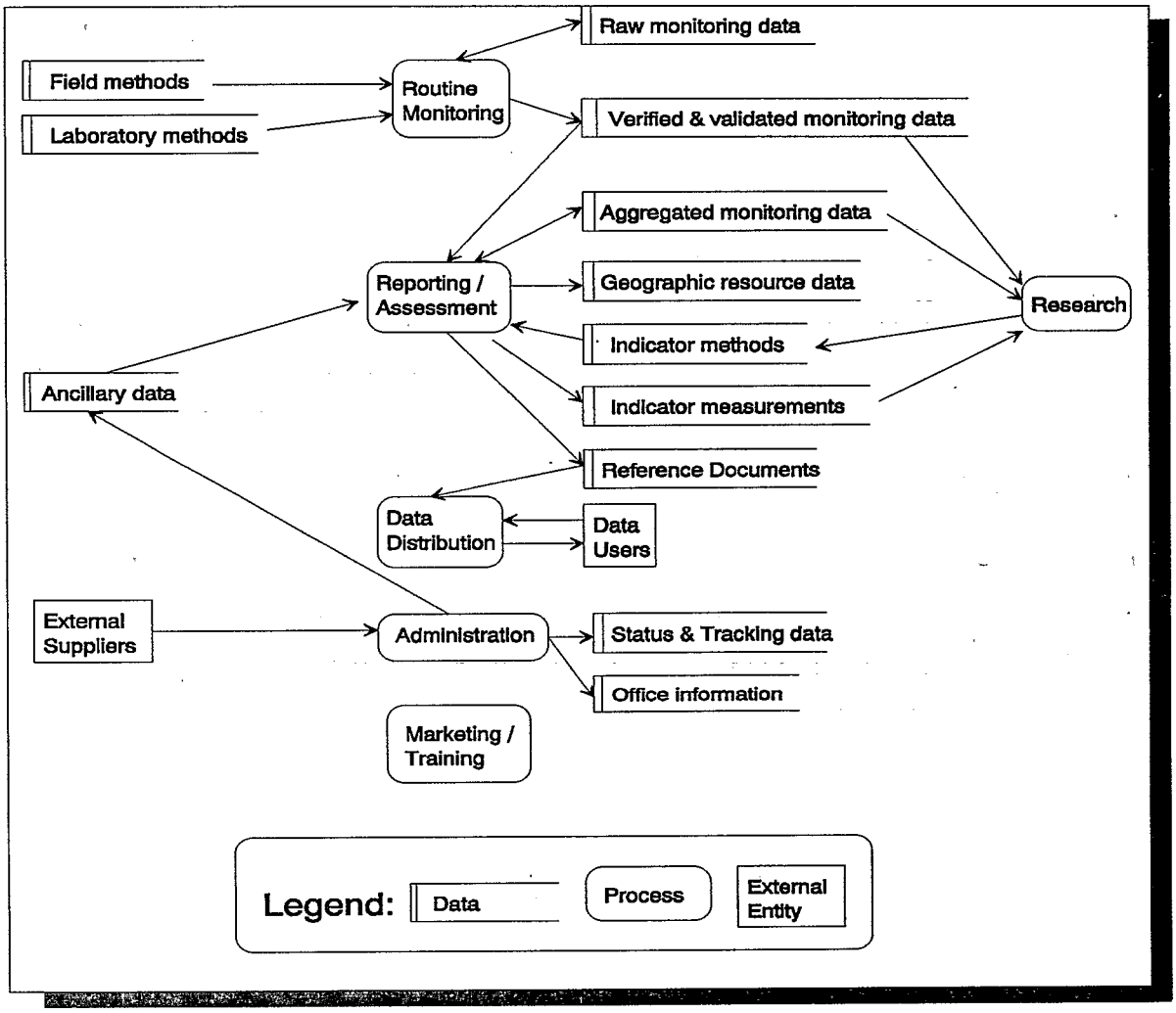

Figure 5.3.2.1. Major EMAP functions.

5.3.3 Network Architecture

Portions of EMAP's technology architecture, discussed in the following section, will be geographically dispersed. In order to accomplish this, a fast and dependable communications network will be required to link the numerous nodes. This network is depicted in Figure 5.3.3.1. As can be seen from the figure, the network architecture is in the form of "node-link-node." EMAP Central and each Resource Group is a node, and the link between them is the communications mechanism, such as a T1 trunk.

EMAP IM is composed of Resource Groups, Coordinating Groups, and a Central Group that are interdependent from an IM perspective. The EMAP IM architecture will be a fully distributed processing system with a peer-to-peer relationship between distributed nodes. Information and processing that may be of use to multiple systems or groups can be housed on the "central" node. It is important to note that this central node will be acting as a peer in the larger system, not as a controlling or centralized hub.

Because of their limited requirements, the Coordinating Groups will not have separate nodes in the EMAP IM system. Instead, Coordinating Groups will use the computing resources of the Central site. This is appropriate because the data that Coordinating Groups use and the crosscutting functions they perform need to be shared among Resource Groups from a corporate perspective.

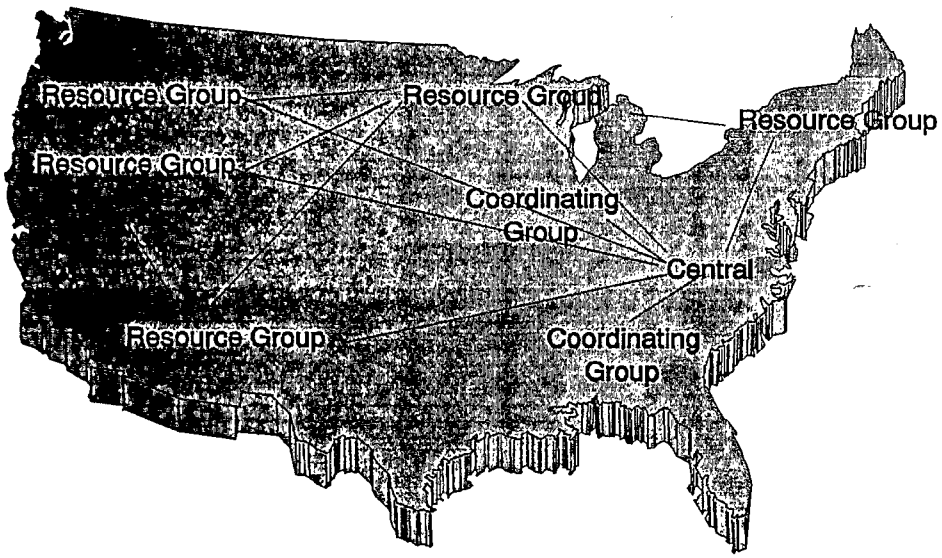

Figure 5.3.3.1 EMAP IM network.

5.3.3.1 Technology Architecture

In order to accommodate rapid and continuous changes in technology, an EMAP IM technology architecture will be developed. Key characteristics of this architecture will include:

- **Open Systems Preference.** EMAP will use industry-standard non-proprietary hardware or software solutions wherever possible. There will be a strong focus on interoperability between various components;

- **Client/Server Approach.** The geographically distributed nature of EMAP IM is ideally suited for a decentralized distributed client/server style of computing. This approach will facilitate the use of large, fast super-minicomputers (such as VAXs and Unix servers) for storing/retrieving data and graphical workstations (for example, Unix workstations or PCs) for interactive analysis and display of data;
- **Multi-Layer Architecture with Interchangeable Components.** Because the EMAP community has large and diverse needs and because change is inevitable and continuous, the EMAP IM architecture will be separated into multiple layers with well-defined interfaces with the respective components. The goal is that as better, faster, or cheaper components become available, they can be included with a minimum of cost, disruption, and perturbation to the overall system; and
- **Integrated User Interface.** To maximize productivity and ease of use, a user interface will be constructed that allows users to access a wide variety of tools in a common manner.

The EMAP architecture must be sufficiently robust to provide access to widely distributed users and information sources. At the same time, it must provide flexibility to attach new tools to the environment so that users can continually enhance their processing capabilities. Figure 5.3.3.2 provides a view of a seven layer model depicting the EMAP architecture. It should be noted that the layers of this architecture are dispersed and may be standalone systems, not *hardwired* or *stovepiped* as the figure may imply.

Forming the foundation of the architecture is the canonical data layer. It is comprised of databases, data sets and documents, and can be independently accessed by any of the tools in the tool set layer. Above the data layer is the Virtual Repository layer. The components of the Virtual Repository are the Data Dictionary, Configuration Manager, Model Manager, CASE tools, Directory and Catalog. These components contain information describing the data that are available, and where they may be found.

The security layer provides the means for EMAP to ensure the long-term integrity of its collection by controlling access. EMAP data must be protected from accidental or malicious damage, and confidentiality of data must be preserved. In order to access EMAP Oracle databases, users must obtain an identification code to allow them entry into the EMAP system. This System Security will be provided through the computer system platforms, and through the varying levels of the Oracle RDBMS security system. Other EMAP data will be available for anonymous, public access through systems such as Wide Area Information Servers (WAIS).

The communications and user access layer is the means by which data may be entered and retrieved from the data layer. Note that the software back plane connects at this level and that access is provided to users of the back plane through the directory

and catalog components as explained in section 5.3.1. Users entering through the tool set layer have the option of directory and catalog access, or through the flexible set of tools that are provided to the EMAP user community to facilitate their use of the data. The inter-process communication layer allows the individual tools to communicate by passing data from one to another for processing. The user interface layer provides easy invocation of the various tools available as part of the EMAP suite, and others available on the client platform.

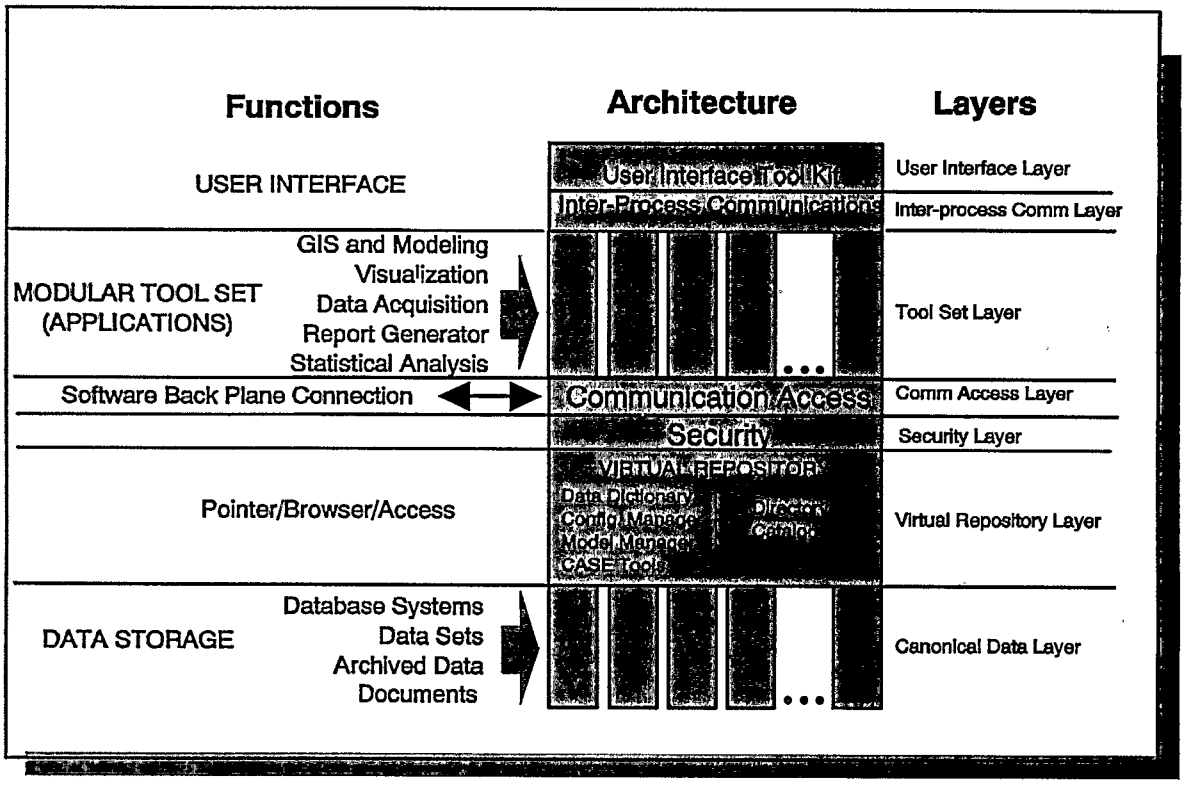

Figure 5.3.3.2. EMAP IM architecture.

5.3.3.2 Technology Architecture Components

Key components of this architecture are listed in Table 5.3.3.1.

Component Name	Component Description	Component Role	Component Examples
User Interface	"Control panel" that allows users to access a wide variety of tools in a common manner.	Provides mechanisms for developing consistent interfaces for a variety of tools.	User Interface Management Systems [70-74], X-windows or windows application.
Inter-Process Communications Layer	Conduit for information to pass from one tool to another, thereby enhancing capabilities of the system and allowing users to analyze information more completely.	Provides tool-to-tool interchange protocols.	DOS, UNIX operating systems.
Tools Layer	Diverse and growing set of information manipulation, data acquisition, and access tools.	Acquires, manipulates, manages, evaluates, and presents information in a wide variety of forms.	Visualization Tools: • Graphics • GIS • Analysis Tools Models • Statistical Packages Hypermedia Systems Data Loggers Database Reporting Systems
Communications and User Access Layer	A tool that provides access to the wide and local area network of EMAP. Users and systems use these communications and access mechanisms to reach information at other locations. A software backplane connection provides a framework for attaching different packages without disturbing applications already connected.	Provides a path between the different systems that comprise EMAP. Allows many physical implementations of architectural components to be connected to facilitate sharing of information and analysis capabilities.	EPA X.25 network, TCP/IP, Internet Dedicated T1, T2, T3 Modems & phone access control cards

Table 5.3.3.1. Technology Architectural Components

Component Name	Component Description	Component Role	Component Examples
Security Layer	Policies, procedures, and technologies to help ensure that information is accessible only to proper users. Integrity of information and systems will ensure that information is not altered inadvertently and that any changes to information will be documented.	Protects system, data, and information assets of EMAP.	Oracle levels of security. Data Processes (transaction committal) Systems.
Virtual Repository Layer	Model Manager - a tool that stores models such as entity-relationship diagrams; data flow diagrams, function hierarchies, and database schemas.	Provides information about entities and attributes of a database model, and what their relationships are.	Infospan Repository System.
	Data Dictionary - contains descriptions, formats, and other basic information for items in a database.	Aids in modeling, developing, and describing the tables and columns stored in a database.	Oracle CASE Dictionary.
	CASE Tools - Computer-Aided Software Engineering.	Provides access and browsing of the encyclopedia and data dictionary.	Oracle CASE*Designer, CASE*Exchange, CASE*Method.
	Directory - a uniform set of descriptions of data sets, data sources, and data set catalogs with pointers to additional details.[48]	Provides the means to index and track data sets, and provides summaries of these data sets.	EMAP Oracle-based directory system.
	Catalog - a uniform set of detailed descriptions of a number of data sets and related entities, containing information suitable for making a determination of the nature of each data set and its potential usefulness for a specific application.[48]	Locates and provides access to detailed documentation on data sets.	WAIS index linked to textual and graphical files.
Data Layer	A diverse and growing set of data/information . This information is stored in various databases and data sets.	Provides organized data for access.	Oracle Databases, SAS Data Sets, CD ROM, ASCII files, Documents.

Table 5.3.3.1 (continued)

5.3.3.3 Architecture Standards

The EMAP technical architecture will use a standards-based open-systems approach. Standards that have been considered[49] for the POC are:

Data Standards

- Data/Metadata Attributes
  - Metadata standard – guidelines for the types of metadata that will be collected and managed.
  - Mandatory data elements for Resource Group data – descriptions of the mandatory types of data that will be collected by different monitoring efforts.
  - Cardinality rules – specify the number of times an occurrence of a data entity can participate in a relationship with another data entity. Cardinality rules allow database integrity to be enforced.
  - Standard for directory structures using directory interchange formats (DIFs).[50]
  - Repository standard – must comply with FIPS-156.
- Naming Conventions
  - Data element attributes – specification of a standard set of data element attributes and their allowable value ranges that is independent of the application usage. This is essential to describing a data element completely for use in a variety of functions.
  - Data element naming conventions – specifications for data element names, descriptive names, and alternate names.
- Locational Data
  - EPA locational data policy which specifies accuracy and resolution of sampling location. EPA policy complies with Federal Interagency Coordinating Committee for Digital Cartography (FICCDC) standards.
  - Spatial metadata standard as defined by the FICCDC.
  - Spatial Data Transfer Standard (SDTS) for exchanging spatial data. SDTS is also a Federal Information Processing Standard (FIPS) standard (FIPS 173).
- Data Codes
  - Taxonomic code – provides a uniform way to identify taxa across Resource Groups that facilitates data integration. EPA currently has a taxonomic standard that is used by a number of other Federal agencies.
  - Tier 3 hexagon identification scheme – identification that correlates sampling location with EMAP Hexagonal sampling grid.

- CAS registry number – provides consistent and unambiguous identification of chemicals as mandated by EPA Order 2180.1.

Network Standards

- Protocols
  - Transmission Control Protocol (TCP) – de facto network standard used by many Federal agencies, universities, and the commercial sector.
  - Government Open Systems Interconnection Profile (GOSIP) – a NIST sponsored standard for network interconnection. GOSIP has been recommended by FIPS 146-1.

Technology Standards

- Database
  - Relational Database Management Systems standards.
  - Structured Query Language (SQL) – guides the method of interactive and programmatic interaction with a database.
  - Database interoperability/remote data access (RDA) – standards for exchanging data between database systems and for accessing data on remote networked computers. These standards are currently being defined by the ANSI X3H2 Committee.
- Computer Aided Software Engineering (CASE)
  - CASE interoperability – definition of a common protocol for exchanging messages between CASE tools and associated repositories and databases. These standards are being defined by the ANSI X3H6 Committee.
- Software Portability
  - Operating system portability (POSIX) – allows programs to run on diverse hardware and software platforms.
- User Interface
  - Graphic user interface style – defines a common look-and-feel for human-computer graphical interfaces. EPA has published a set of standards for agency systems.
  - Data entry user interface style – defines a common look-and-feel for text-based human-computer interfaces. EPA has published a set of standards for agency systems.
- Scientific and Technical Information Retrieval
  - Standard Generalized Markup Language (SGML) – a document interchange standard.
  - Search and Retrieval Standards (ANSI Z39.50) – common protocol for searching and retrieving text-based information.

Wide Area Information Server (WAIS) - information retrieval protocol and software for distributed information retrieval systems.

5.3.4 People Architecture

To fully determine and understand the user requirements for EMAP, an Enterprise-wide people architecture will be produced using the Framework described in Section 5.3.2. The organization design challenge has to do with the allocation of work and the structure of authority and responsibility. Therefore the structure of the people architecture is in terms of "people-work-people". This architecture will be used to ensure that EMAP's personnel organization is considered during the systems life cycle. The people architecture will also be used to facilitate communication and information exchange within EMAP. Organizational charts, as shown in Figures 2.3.2 and 6.4.1 and Table 6.4.1, will be used to determine roles of the various EMAP offices as well as the work products and deliverables. The people architecture will be specifically considered in the development of the user interface and system security architecture layers.

5.3.5 Time Architecture

Along with the other Framework architectures, the time architecture is used to determine system requirements. The time architecture, as guided by the "event-cycle-event" process of the Framework, will ensure that the EMAP IM system can accommodate the phasing of requirements of users' ADP loads. Understanding these requirements early in the life cycle will aid in the management of EMAP IM assets.

5.3.6 Motivation Architecture

The motivation architecture in the Framework is comprised of descriptive representations that depict "why" the Enterprise is being developed. This architecture is expressed in terms of an "ends-means-ends" model. For EMAP, the ends are the objectives, or goals, as described in Section 2. For example, the purpose for integrating scientific data sets into structured products must be explicit in order to design the information management system. The strategies, or means, employed to meet these goals will be defined in documentation prepared at the appropriate motivation architecture level. This Strategic Plan addresses why IM is important within the "scope" row of the Framework.

Section 6

PROJECT MANAGEMENT APPROACH

6.1 Introduction

The EMAP IM Project Management Approach provides the framework for rapid, continual, and measurable development of information systems for EMAP IM. The principal components of the EMAP IM management approach are:

- Project management model, and
- Functional organizations which will implement the model.

The project management approach must address the objectives as they are described in Section 2, *Background.* It must enable the development of systems that satisfy the requirements as described in Section 3, *Users and Requirements,* and since this program is long term, the proposed management approach must follow certain *Guiding Principles,* as outlined in Section 4. The IM project management approach is designed to accommodate the *Technical Approach* as discussed in Section 5.

6.2 Project Management Challenges

The success of any Enterprise, including the EMAP program, is strategically dependent on the success of information management.[51] The project management challenges facing the EMAP IM management team include the following:

- IM systems must be designed to support continuously changing user requirements. The types of data monitored and the number and kinds of information products that result from this data will grow dramatically over time;

- IM systems must be open standards-based and extensible in order to evolve over time and include: new technology advances, enhancements to existing systems, and leverage with other internal EPA development efforts;
- EMAP IM must be prepared to provide information products that span a wide spectrum of users including scientists, policy makers, and the public at local, State, and National levels;
- Field monitoring and analysis activities must be ongoing and, once obtained, monitoring and analysis results must never be lost or rendered unusable; and
- Monitoring and analysis activities will expand beyond national boundaries to provide global evaluation.

6.3 Project Management Model

To meet these challenges and to provide effective coordination and communications across the EMAP community, a systems engineering project management model is being used by EMAP IM in the development, management, and implementation of IM systems. Following this model will increase management confidence in project schedules through responsive solutions. Resulting benefits include:

- Ability to cope with complexity through iterative refinement of architecture and requirements;
- Early validation of concepts and recognition and elimination of unnecessary functionality;
- Emphasis on well designed interfaces between system components; and
- Integration of quality assurance and product development activities and increased product reliability, maintainability, and usability.

Ultimately, the success of the systems engineering approach will depend on correctly structuring the management of EMAP IM to address the functional activities of EMAP IM systems development and operation. The designation and use of specialized functional teams will facilitate the implementation process.

6.3.1 Information Systems Management Model

The EMAP IM information systems management model is depicted in Figure 6.3.1.1. It is a high-level adaptation of an IBM Corporation information systems management process model.[4] This model was chosen since it accommodates dispersed management functions such as exists with EMAP. There are three levels of the process: strategic, tactical, and operational. (In the context of the Zachman Framework, the strategic level encompasses scope and the enterprise model; the tactical level encompasses the technology model and components; and the operational level encompasses functional systems. See Figure 7.5.1.) These levels are then sectioned into development, management, and service missions. The

evolutionary technical approach of EMAP IM engages all levels and missions virtually simultaneously since operational implementation must continue to occur within the Resource Groups while EMAP IM strategic and tactical events are still evolving. Because of operational requirements, EMAP IM is an aggregation of selected systems engineering methodologies that are best suited to EMAP needs. Discussion of discrete tasks for each level and mission follows Figure 6.3.1.1.

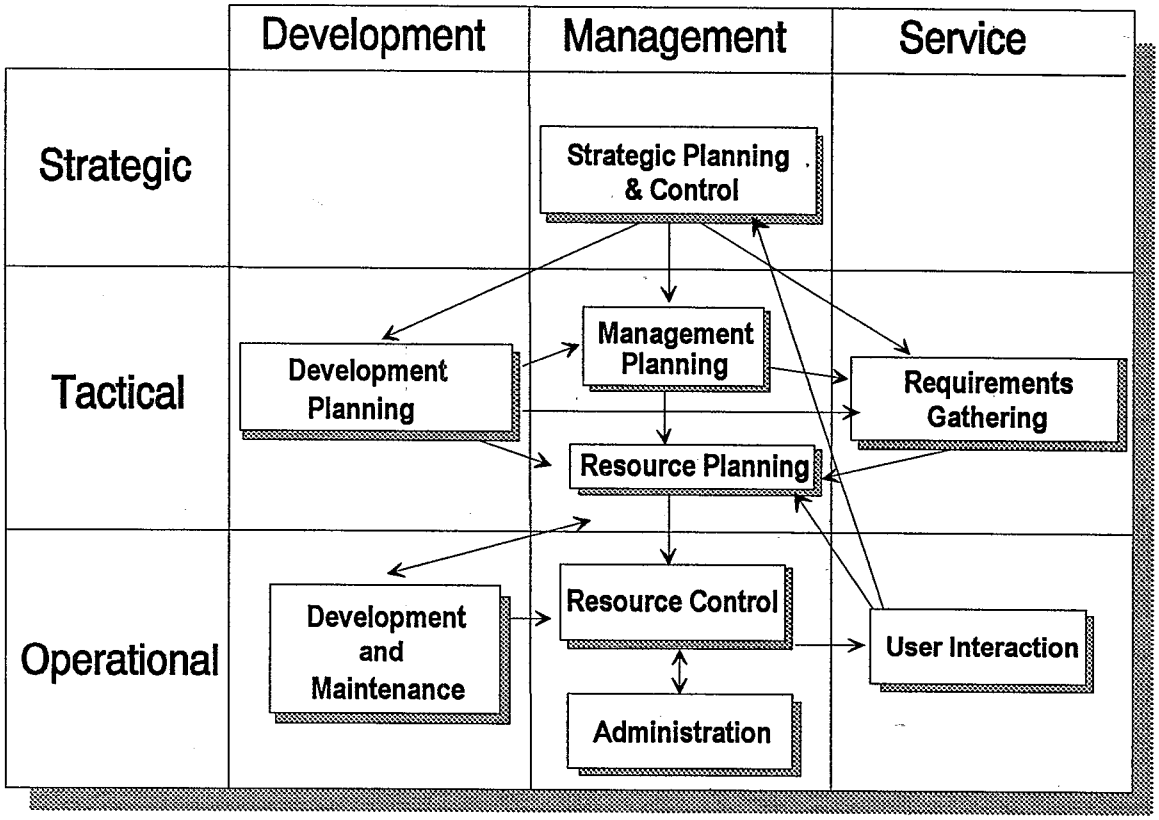

Figure 6.3.1.1. EMAP IM information systems management model.

Strategic Planning and Control. This strategic process defines EMAP IM Enterprise expectations of the information system functions through the Strategic Plan period and how they will be met. This Plan defines the Enterprise mission and objectives, policies, and requirements.

Tactical Development Planning. The tactical development processes translate the strategic goals and direction into an application plan that encompasses the key functions of the EMAP IM Enterprise. A data plan is then derived from the application plan. Using the application and data plans and user requirements, a systems plan for hardware, software, and the network is developed. These plans then translate into a manageable, documented plan.

Tactical Management Planning. Using strategic guidance and an assessment of the existing management systems, this process defines a prioritized management plan to improve the management system. Quality control and quality assurance guidance must be provided. For EMAP IM, a Proof of Concept (POC) is the first project management activity. The POC will define technical feasibility, better define resources for the Enterprise, and justify and prioritize the management systems plan.

Tactical Resource Planning. Resource planning is required to verify and validate IM resources for EMAP Central, Resource and Coordinating Groups, and Cooperative Partners. Considering users requirements, the appropriate number and qualification of personnel must be determined and budgets then developed. Tactical management of the project plan will then occur on an interactive feedback basis.

Tactical Requirements Gathering. This process evaluates characteristics of EMAP IM users and gathers functional and system requirements. Configuration control, security, and feedback processes must be considered at this level of the information system management process.

Operational Development and Maintenance. As described in this Plan, this process has two components: a control or oversight function and an applications and software development function. Using the tactical plan, this process defines the project scope, leadership, and user involvement necessary to ensure successful implementation. A detailed project plan is developed including objectives, resources, time frame, tasks, organization, and deliverables. Applications, software, and documentation are developed and modified as required by the users. Project reviews are conducted with *Management* and *Service* users.

Operational Resource Control. This part of the process determines change requests and adjusts budgeting and tactical planning accordingly. Using change information, this process builds and manages inventories of all EMAP IM resources.

Administration. Such administrative activities as financial administration, education, and training are managed at this level. Education activities include educating information management professionals about science issues and educating scientists about information management issues. Also, this process includes maintaining contractual agreements for project work and other supportive efforts.

Operational User Interaction. This process validates user information requirements. With EMAP IM, this processed data and information is translated into monitoring and assessment reports and annual summaries. Requests from other users would be considered as well. Changes to the information system process are provided to the strategic and tactical planning process as a feedback mechanism for adjusting to changing requirements and technology.

6.4 EMAP IM Functional Organization

Figure 6.4.1 shows the functional organization for EMAP IM systems development. Responsibilities and interfaces for each of the primary functional entities as they participate in EMAP IM systems development are summarized in Table 6.4.1.

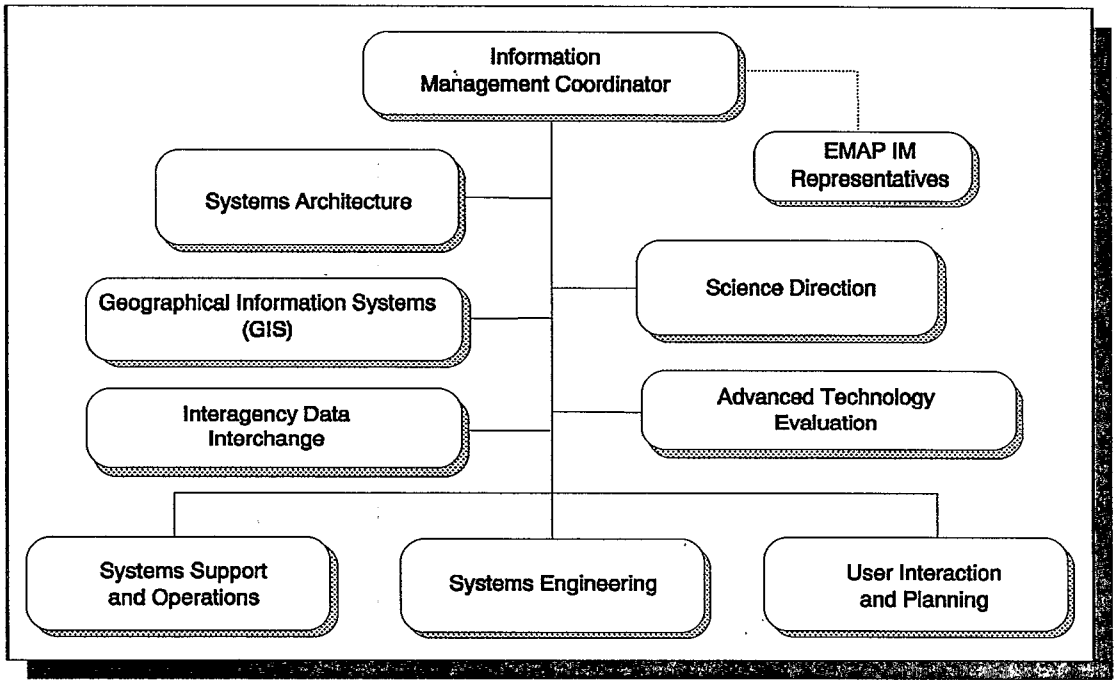

Figure 6.4.1. EMAP functional organization for IM systems development.

Function	Responsibilities
Coordinator of IM	• IM direction, organization and staffing • Assure adequate resources for development, communications and operations • Strategic and tactical planning • Project performance evaluation
Science Direction	• Annual scientific review with external scientific peer reviewers from other agencies, universities and national laboratories • Provide technical advise and direction on information management issues
Systems Architecture	• Define and enforce systems architecture standards • Evaluate standards and encourage development if not available • Maintain architecture and standards
Geographical Information Systems	• Develop GIS base maps and coverages • Liaison with GIS groups in other agencies • Investigate and test new GIS technologies
User Interaction and Planning	• User requirements definition • User training and manuals • User involvement in IM activities • System needs and requirements feedback • Planning for new functionality
Interagency Data Interchange	• Liaison to IRM groups in other agencies • Develop agreements (MOUs) with other agencies • External data exchange protocol standards • Directory and catalog design
Systems Engineering	• Detailed engineering specifications • System engineering studies • System development and documentation • System testing • Systems deployment
Systems Support and Operations	• Ensuring operations of central and distributed database systems • Data administration and inventory management • Maintain access and security systems • IM product release management • Standards enforcement • Configuration management/change control • HW/SW procurement planning • Quality assurance of IM products
Advanced Technology Evaluation	• Information sciences technology evaluations • Studies on advanced technologies such as: Database management systems; data and metadata management; and configuration management • Prototype advanced technology alternatives • Tool and data integration prototyping such as: Visualization techniques and multimedia reporting framework

Table 6.4.1. Responsibilities of EMAP IM Functions

6.5 EMAP IM System Management Functions

The Resource and Coordinating Groups will employ information management staff and support contractors to operate data collection, data definition, data librarians, GIS, and analysis and information translations and flows. The information system management responsibilities for EMAP IM functions are illustrated in Figure 6.5.1.

	Development	Management	Service
Strategic		• IM Coordinator • Coordinating Groups • Science Direction	
Tactical	• User Interaction & Planning • Resource Groups	• IM Coordinator • System Architecture • Technology Evaluation • Coordinating Groups • Resource Groups	• User Interaction & Planning • Interagency Data Interchange • Resource Groups
Operational	• GIS • Systems Engineering • System Support & Operations • Coordinating Groups • Resource Groups	• IM Coordinator • Coordinating Groups • Resource Groups	• User Interaction & Planning • Coordinating Groups • Resource Groups

Figure 6.5.1. EMAP IM information systems management functions.

6.6 EPA Functional Interactions

A number of functional interactions are required to develop EMAP IM systems that operate within the EPA information systems environment. Some of these interactions are explained below.

Bay City

Numerous projects and activities are being developed in EPA, other agencies, industry, and academia with which EMAP IM systems must maintain an active relationship.[52-57] Interactions with these activities will enable technology transfer among all participants. One such project is the U.S. Global Climate Research Program (USGCRP), and sub-projects within USGCRP that are of direct interest. The EPA is participating partly through the EPA Bay City Facility.

The Bay City facility is proposed as the information sharing vehicle between EPA's EMAP Program and the Interagency Global Change Program. This facility will be a super computer center focusing on ecological and other environmental modeling. Their systems will be designed to enhance the understanding of ecological resource trends. EMAP information will be an important input to these systems and should

provide a baseline for understanding these trends. The interactions through Bay City are:

- Collaboration and data link to the interagency Global Change Program and the Earth Observing System/Distributed Information System (EOS/DIS);
- Coordinating point for the several components of the EPA's Geographic Initiatives; and
- Establishing a linkage to ecosystem theory.

The Bay City facility, as an Affiliated Data Center, will be patterned after the NASA Earth Observing System (EOS) Distributed Active Archive Centers (DAAC). The Global Change program is mandated to enable interoperability and data exchange between agency data systems (including EPA). In addition, this Center will form direct linkages to the Interagency Working Group on Data Management for Global Change.

It is proposed that an EPA Data Center be established at the Bay City facility to coordinate interactions between EMAP and the Global Change Program. Also, the Spatial Data Center would include results from the EPA Geographic Initiatives including the Great Lakes Ecological Process Pilot, Regional Oxidant Model, or the Ecoregions thematic map. In addition, this Center would provide access to external spatial data, such as Census TIGER, Digital Elevation Models, World Cartographic Database, and National Wetlands Inventory.

The Bay City facility, with its current focus on super computing applications will also provide the linkages to theoretical environmental modeling. Data from EMAP IM systems will be critical for developing, calibrating, and establishing a baseline for ecological resource trend predictions.

EPA EnviroFacts and Gateway

EnviroFacts is an EPA project to develop a central data repository to support cross-media environmental management and decision making. This repository, using the information warehouse concept for long-term information storage, will have access to key data extracted from existing EPA program databases, new data collected as part of EMAP, metadata, and an index and pointers to other data not contained in the repository (e.g., other Federal databases and environmental data).

The repository will exist in a standard, centralized data processing environment. It will have powerful access and analytical capabilities and will reflect EPA data and technology standards. The Gateway project will provide a common user interface to this repository. EMAP IM interactions and coordination with EnviroFacts and Gateway will continue as the projects develop.

EPA Center for Environmental Statistics (CES)

The objectives of the CES are to:

- Provide a central environmental data source;
- Prepare and publish analyses of environmental data to support policy making, program evaluation, public information, and education;
- Routinely prepare integrated statistical assessments of national environmental conditions and trends; and
- Help public officials and citizens find environmental information.

This is an emerging initiative within EPA that can use EMAP data and potentially have direct linkages in its routine statistical reporting capabilities.

STORET Modernization

The Office of Water within EPA developed an information management system called STORET in the early 1960s to accommodate water sampling information collected as part of national monitoring events. This system provided for the standardization of collection, analysis, and storage methods of the sample information. STORET is presently being modernized to:

- Standardize systems software/tools;
- Integrate with data from other systems;
- Allow for the quality of data; and
- Simplify system use.

The modernization effort commenced in 1991, and the new system will be fully implemented in 1997. EMAP IM will coordinate with the modernization efforts of STORET to share common data and information and preclude duplication of effort.

EPA Geographic Initiatives

It is important that EMAP IM coordinate with the Resource Groups associated with specific geographical initiatives. At present, this interaction is expected with EMAP Great Lakes on the Canadian border and with EMAP Estuaries in the Gulf of Mexico.

Section 7

IMPLEMENTATION PLAN

7.1 Introduction

The Environmental Monitoring and Assessment Program (EMAP) Information Management (IM) implementation plan defines the details for implementing EMAP IM systems following the evolutionary approach described in Section 5, and using the information systems management model discussed in Section 6. This section provides the high-level implementation process, specifically:

- What is to be implemented;
- How it is to be implemented; and
- Who will conduct the implementation within the EMAP functional organization.

Also included is an implementation schedule for each of the three implementation processes: *Proof-of-Concept* (POC), *Technology Transfer*, and *Enterprise Implementation.*

7.2 EMAP Enterprise

The EMAP Enterprise, when fully developed, will include Resource Groups, Coordinating Groups, EPA and non-EPA related programs, and administrative activities that support the program. Collectively, EMAP will be an enterprise made up of these user groups, each of which have differing but complimentary missions to perform in order for EMAP to accomplish its objectives. But since EMAP is and will continue to be an evolving program with diverse user requirements, implementing the IM aspects of EMAP requires a stratified approach with a map to describe what is to be implemented and the priority of implementation efforts. Figure 7.2.1 is that map.

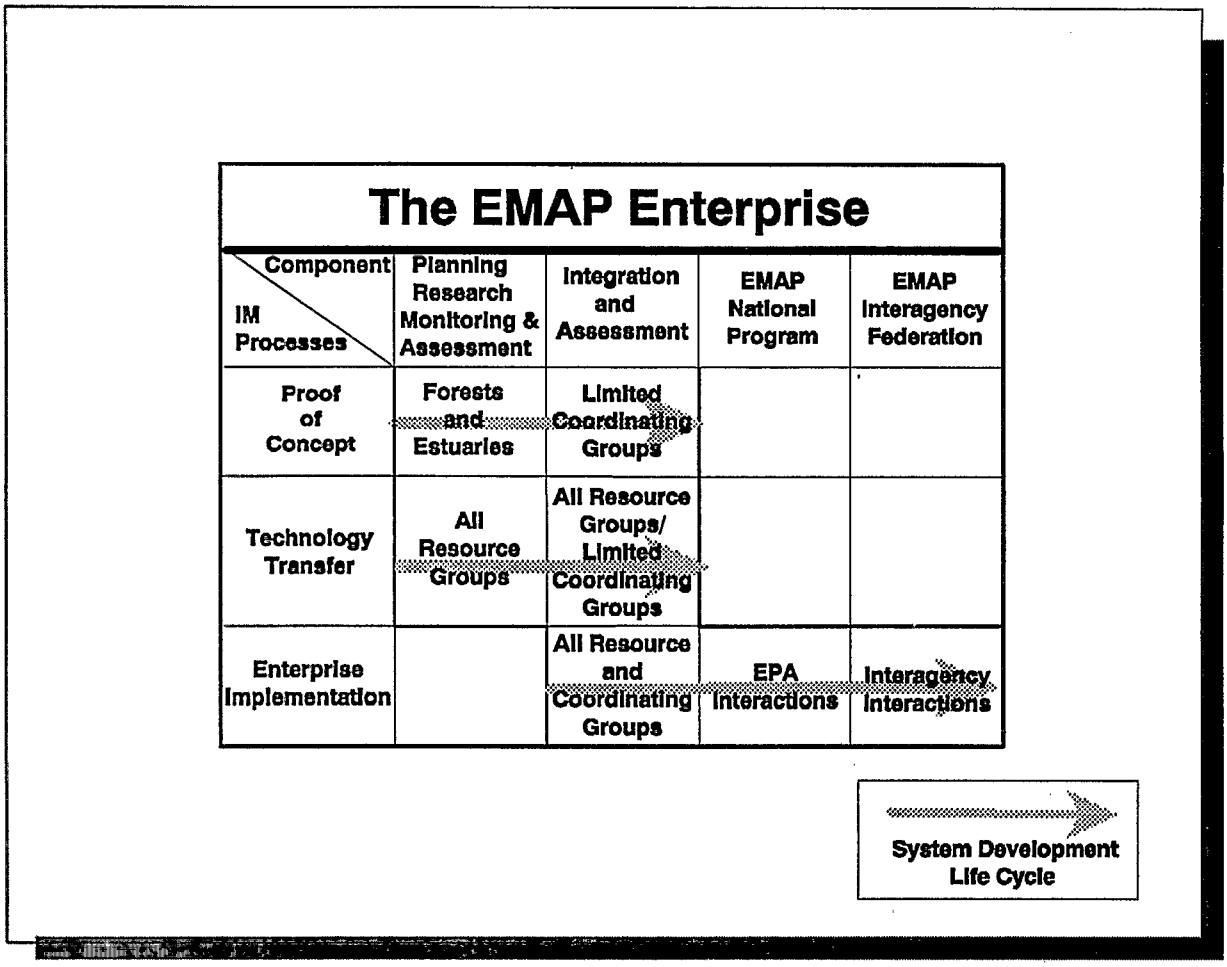

Figure 7.2.1. The EMAP enterprise.

7.3 Implementation Components

The size and diversity of the collective EMAP user constituency, and the evolving nature of EMAP requires the development and implementation of a strategic systems engineering approach stratified by groups comprised of similar users, with similar missions, and in similar stages of development. These groups are separated into four components shown in Figure 7.2.1.

7.3.1 Planning, Research, Monitoring, and Assessment

The Planning, Research, Monitoring, and Assessment component includes planning, indicator research, collection and analysis of data, and distribution of data and information related to an ecological area of responsibility. This component is primarily focused on the Resource Group users who will have operational systems after completion of the Technology Transfer process.

7.3.2 Integration and Assessment

The Integration and Assessment component will provide direction on the composition of EMAP information products that are derived from data available within individual resources as well as information products derived from multiple resource groups. The Enterprise Implementation process is complete for the Integration and Assessment Component when the full needs of all Resource and Coordinating Groups are met in producing the required information products.

7.3.3 EMAP National Program within EPA

The EMAP National Program component within EPA defines the relationship of EMAP to other relevant EPA programs. These programs include EPA Geographic Initiatives such as the Great Lakes National Program, the Gulf of Mexico Program, the STORET Modernization within the Office of Water, the Gateway/EnviroFacts program within the Office of Information Resource Management, and others. This component also includes connections to appropriate EPA administrative programs.

7.3.4 EMAP as part of an Interagency Federation

EMAP as part of an Interagency Federation defines the relationship of EMAP to relevant programs outside of EPA that are not already included in other EMAP enterprise components by virtue of their cooperative partner status. Examples of the type of programs targeted are NASA's EOS/DIS and the Global Change Research Program, both of which have no direct and single counterpart among the Resource Groups.

7.4 Implementation Processes

Because of the size and varying degrees of maturity with each user group component, EMAP IM has stratified development and implementation into three discrete process phases depicted in Figure 7.2.1. Each process represents an execution of the systems development life cycle, the result of which is a set of operational information management systems that meet the needs of the users shown in the map cells.

7.4.1 Proof-of-Concept

The Proof-of-Concept (POC) process is the pilot phase of the Enterprise. This process identifies and applies standards, policies, tools, and procedures to two Resource Groups and selected Coordinating Group participants. A goal of the POC process is to construct databases for the Forest and Estuaries Resource Groups. The key objectives are to:

- Develop operational systems for Forests and Estuaries;

- Build data models to support monitoring data collected by the Forests and Estuaries Resource Groups;
- Implement the data models at Las Vegas, Narragansett, and EMAP Central to demonstrate the capability of the Enterprise;
- Demonstrate the ability to transfer data between users at different geographical locations;
- Develop corporate, warehouse, and archival databases at EMAP Central; and
- Identify infrastructure requirements to support future efforts.

The POC involves linking data, metadata, and information among the two EPA Resource Groups and EMAP Central. Scientific data and metadata will be located at the Resource Groups with corporate data, metadata, and information located at EMAP Central. The success of the POC hinges on the demonstrated communication among these users across the POC system architecture. Communication depends on sufficient standardization and integration of system components within the architecture. These components include common protocols, common user interfaces, compatible data management systems, common data exchange mechanisms, common tools, etc., to share EMAP information among multiple and dispersed users.

7.4.2 Technology Transfer

The Technology Transfer process involves sharing the technology validated during the POC with the remaining Resource Groups and an increasing number of Coordinating Groups. The objectives of this process are to:

- Support each Resource Group with their development of an IM infrastructure required for their planning, research, monitoring, and analysis functions; and
- Carry on the process initiated with the POC of developing the IM capability needed to meet all EMAP program objectives.

The Technology Transfer process will be conducted in two phases—requirements and implementation. The requirements phase will consist of a review of existing documentation, creation of models as depicted in the Zachman Framework for each Resource Group, and building integrated EMAP models that show common processes and data shared across Resource Groups. Each Resource Group should have an operational system at the end of this process. Additionally, EMAP Central should have realized an increase in Coordinating Group involvement.

7.4.3 Enterprise Implementation

The Enterprise Implementation process involves remaining EMAP users and U.S. EPA and other agencies as EMAP components expand in scope. Enterprise Implementation is the process that re-engineers the preceding processes in order to fully

integrate and implement a data transfer capability at an enterprise level. EMAP IM enterprise objectives are to:

- Harmonize EMAP IM among all Resource Groups and Coordinating Groups after technology transfer is complete;
- Harmonize EMAP IM with other EPA programs and functions after Integration and Assessment Enterprise Implementation is complete; and then
- Harmonize EMAP IM with Interagency Federation participants outside of EPA such as NASA's EOS/DIS and the Global Change Research Program.

7.5 Management of the Implementation

This section discusses the events that represent the confluence of technical and management methodologies for EMAP IM. Figure 7.5.1.1 represents a combination of highly aggregated management functions from Figure 6.3.1.1 superimposed on the Zachman Framework. The *evolutionary systems engineering approach* is represented by the EMAP process development life cycle applied to each column or architecture within the Framework. The remainder of this section explains implementation events using the information systems management model from Figure 6.3.1.1.

7.5.1 Strategic Planning and Control

The Strategic Planning and Control function provides direction to the entire Enterprise process through plans, reviews, and evaluations of evolving EMAP IM systems. This activity provides guidance to development activities and provides critical technical review from a programmatic perspective.

Strategic planning has been ongoing since 1992 as versions of this Strategic Plan were reviewed by an ever-broadening group of reviewers. Each version of the plan has been progressively more detailed, responding to feedback from management and users to lay the foundation for activities in subsequent years.

Architecture standards will be defined for the POC and Enterprise Implementation processes. These standards will support all aspects of the systems architecture from the system development life cycle (SDLC) perspective. Architecture planning will closely align with the Information Management/Data Administration (IM/DA) team in the Office of Information Resources Management (OIRM). The architecture team will work closely with the Architectural Management and Planning Branch (AMPB) of the National Data Processing Division (NDPD), EPA Geographic Initiatives, the Gateway/EnviroFacts effort, STORET modernization effort, Global Change program, and the EOS program to identify, develop, and promulgate data, process, and technology standards that assist in integrating IM among these efforts. The architecture team will work closely with the Information Management and Services Division of OIRM

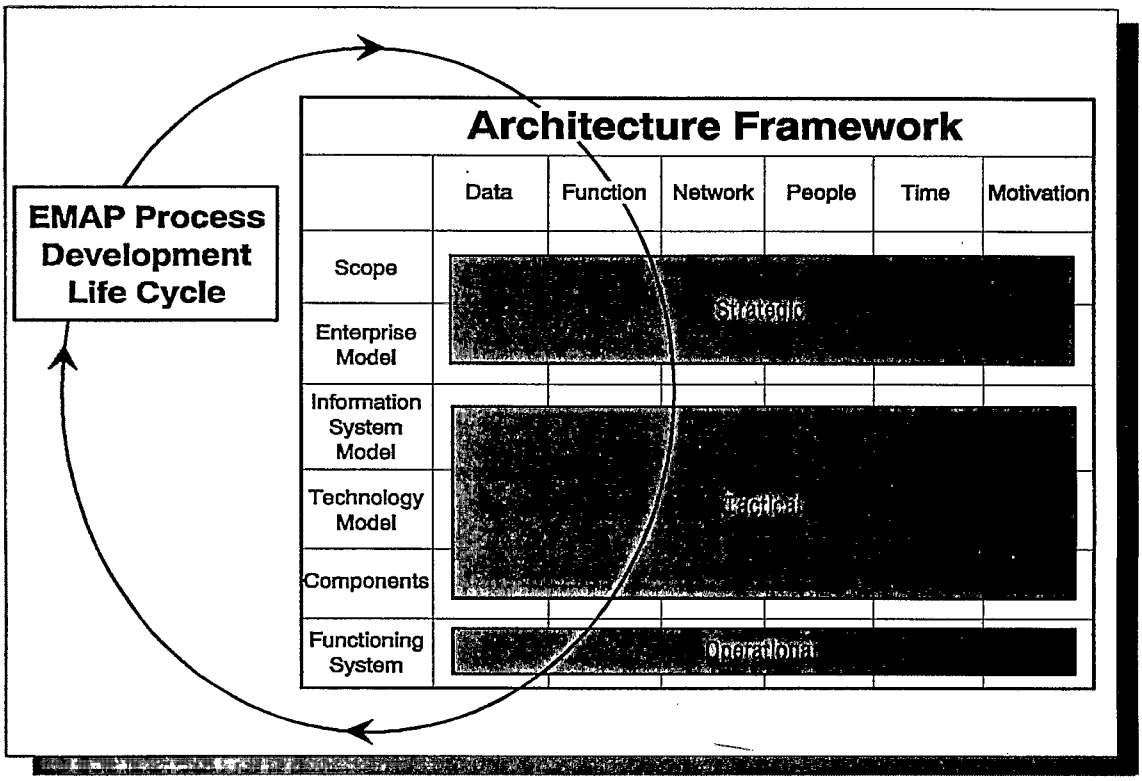

Figure 7.5.1.1. Architecture framework with management levels.

in order to develop and adopt a standard SDLC for EMAP that is compatible with the EPA standard approach.

Another key function of strategic planning is review at technical peer, EPA/IRM, and the Science Advisory Board (SAB) levels. These reviews must take place at scheduled dates relative to project progress and strategic points when critical review can influence direction. The strategic planning activity is the responsibility of the EMAP IM Coordinator and is implemented in conjunction with the Science Direction and Systems Architecture functions. The products of this activity are versions of the Strategic Plan, architecture standards manuals, and management and technical guidance for EMAP IM systems development.

7.5.2 Development Planning

Development planning activities are the responsibility of the User Interaction and Planning function in conjunction with the Resource Groups. This activity will start with the development of several Enterprise models to clarify user expectations of needs within the context and scope of the EMAP Project. These models will be the primary source for user requirements. They will be developed using facilitated team techniques such Joint Application Design (JAD) sessions.

Facilitated teams of users work best when the team members have a common global view of the EMAP program. Requirements from the principal EMAP IM users will be addressed to have a complete model of user expectations. Resource Group users are considered the highest priority for initial formal needs development. This allows initial functional and engineering specifications to be developed prior to full completion of the Enterprise models. These activities will produce a documented understanding of the users expectations such as found in the "Entity Relationship Diagrams for EMAP Resource Groups."[58]

7.5.3 Management Planning

Management planning activities are the responsibility of the IM Coordinator and the information managers of the Resource and Coordinating Groups. Using strategic guidance and an assessment of the existing management system, this process defines a prioritized management plan to improve the management system through a project approach. The System Architecture and Advanced Technology Evaluation functions will support this activity through evaluations and assessments of future architecture standards and hardware, software, and network technology.

A Tactical Plan will be prepared each fiscal year to specify the statements of work for the respective functional support contractors and a Project Management Plan will specify all required tasks to include interrelationships, time lines, and deliverables.

7.5.4 Resource Planning

Resource planning is the responsibility of the IM Coordinator and the Resource and Coordinating Groups. Estimates of cost alternatives for potential technology improvements will be provided in a separate document by the Advanced Technology function. Projects with large information components must usually invest heavily in the early stages of the overall program. This project is no exception. Providing the IM systems prior to large-scale national monitoring ensures that data will be managed properly, be of known quality, and be available for long-term analysis.

Staffing and staff cost estimates are provided in this section in order to plan resources for the period of this Strategic Plan. These estimates are for strategic planning purposes and do not reflect detailed tactical planning activity. Table 7.5.4.1 lists personnel estimates for each of the major activities within the EMAP IM systems development functional organization. These numbers are an aggregation of EPA, Cooperative Partner, and contractor personnel. These activities are supported by the organizational functions discussed in Section 6. Field support, however, includes Resource and Coordinating Group staff and staff from Cooperative Partners who actively collect, manage, and verify ecological resource data.

Activity	FY 93	FY 94	FY 95	FY 96	FY 97
Project Management	2.5	2.5	3.5	3.0	3.0
Science Direction	0.5	0.5	1.0	1.0	1.0
Geographical Information System	1.0	1.0	1.5	1.5	1.5
User Interaction and Planning	6.0	6.0	6.0	8.0	8.0
Advanced Technology Evaluation	—	—	1.0	3.0	3.0
System Architecture	3.0	3.0	4.0	3.0	3.0
Interagency Data Interface	0.5	0.5	1.0	2.0	2.0
Systems Engineering	1.0	1.0	10.0	13.0	8.0
Systems Support and Operation	0.5	0.5	2.0	3.0	3.0
Subtotal	**15.0**	**15.0**	**30.0**	**37.5**	**32.5**
Field Support	27.0	33.0	36.0	42.0	42.0
Total	**42.0**	**48.0**	**66.0**	**79.5**	**74.5**

Table 7.5.4.1. Staff Estimates for EMAP IM Systems Development

Field Support is a matrix-managed activity with funding support provided by the individual Resource and Coordinating Groups as well as the Cooperative Partners. The total budget required for EMAP IM is on the order of $9.0 -12.9M per year for the next 4 years (see Figure 7.5.4.1). Approximately $8.2M was spent in FY 93. This includes field support, EMAP Central support, and hardware and software. These figures are based on an average of $150,000 per person per year for all functions. This budget is expected to support the minimum information systems functionality required for the EMAP IM strategy.

7.5.5 User Requirements Gathering

User requirements definition and design are the responsibilities of the User Interaction Planning and Interagency Data Interchange functions in conjunction with the end users. These users are the Resource Groups, the Coordinating Groups, and those EPA activities such as the laboratories which perform the EMAP planning, research, monitoring, and assessment tasks. EMAP IM must support the development, operations, and maintenance of both central and distributed information nodes throughout the full range of the system life cycle.

A fundamental premise for EMAP IM success is involvement of users at all stages of the development cycle, including architecture, design, and engineering processes. Further, users will assist in the test for completeness of the functional specifications and IM architecture at frequent demonstrations of system functionality to ensure that their expectations are met. Evolving EMAP IM systems must fit within the users' current operational environment or transition plans must be developed.

Requirements gathering is an iterative process and applicable to all implementation processes. Requirements for the POC will be limited to two Resource Groups, and

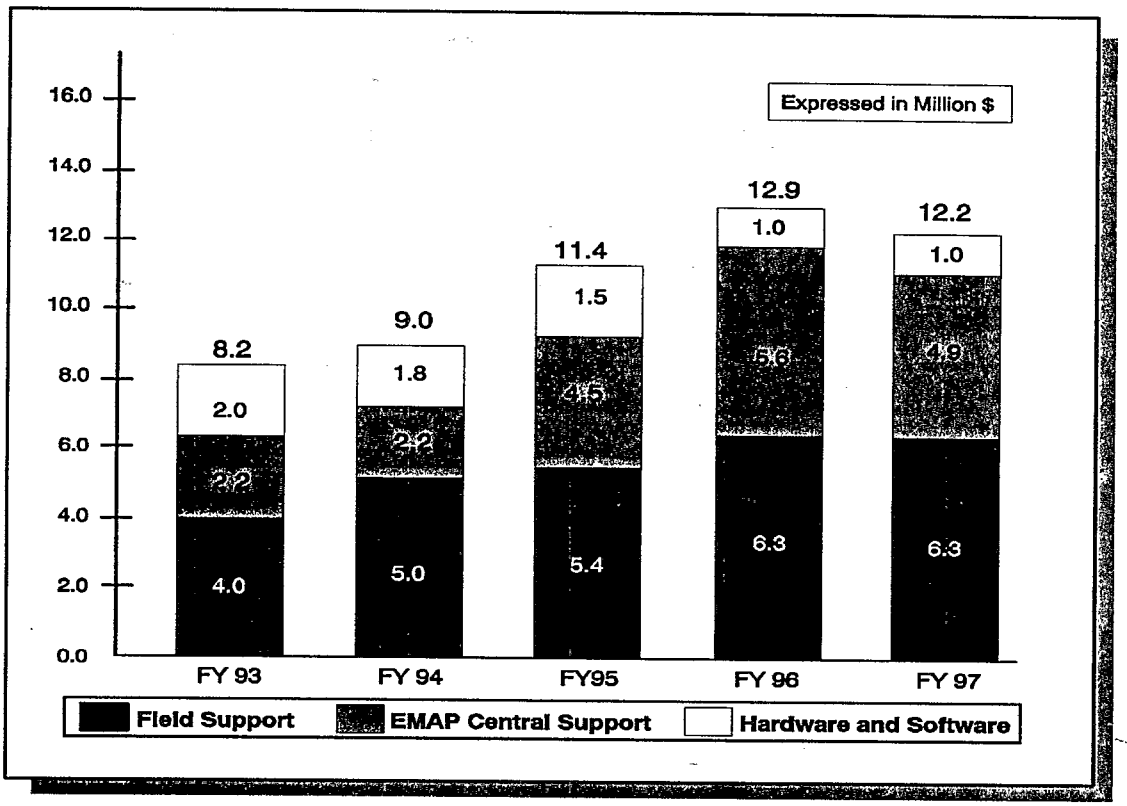

Figure 7.5.4.1. Estimated resource requirements.

Joint Application Design (JAD) sessions will be held when required to define and design data, process, technology, and network requirements. Initial requirements are being limited to the POC; however, additionally requirements gathering will extend through the Technology Transfer process with the other Resource Groups. After the POC technology is transferred to other Resource and Coordinating Group users, additional JAD sessions will plan for the Enterprise process. In addition to JAD sessions, requirements will be verified through user interviews and demonstrations of prototypes at designated IM meetings. Results from the requirements gathering process will include:

- Security and access control requirements;
- Entity relationship diagrams;
- Data and metadata, process, technology, and network requirements;
- Directory and catalog requirements;
- Interface and query requirements;
- Data flow and process flow diagrams;
- Training and implementation planning;

- Archiving and legacy information requirements; and
- Other system development life cycle planning.

7.5.6 Development

Development is the responsibility of the Systems Engineering function in coordination with the users. They will first develop a POC pilot for the central node and assist in the implementation for Forest and Estuaries Resource Groups. The actual databases with hardware and software interfaces and applications will be implemented and tested with the Forests and Estuaries Resource Groups and the Coordinating Groups at EMAP Central. As the technology is validated and transferred to other Resource Groups, the architecture will become increasingly decentralized, or "federated," until all participants become responsible for their own components of the Enterprise.

For the entire evolutionary development process, Systems Engineering will conduct detailed design, coding, testing, and implementation of selected prototypes. This will include developing the physical databases to support prototypes and pilots, code and test prototype application programs, and develop appropriate documentation.

7.5.7 Maintenance and Operations

The maintenance function provides for the ongoing operations of the hardware and software configurations used for processing EMAP's system transactions and is the responsibility of the Systems Support and Operations organizational function. This function will provide the data administration and inventory management of the distributed databases and implement and manage the access and security processes associated with daily operations. Responsibilities include installing change management procedures and coordinating and documenting version control of all EMAP IM software across all client servers. Additional tasks will include installing new versions of EMAP applications and tracking data and product distribution.

Inherent to this function are the quality assurance (QA) measures which must be followed to ensure QA reviews of all systems and application products for the EMAP IM system. This will include walk-throughs of general IM support systems as well as specific EMAP system products. The purpose of these walk-throughs is to ensure that newly installed systems will not adversely affect overall service levels or cause problems with ongoing maintenance of these systems. They will follow EPA-specified procedures for ongoing maintenance, backup, and change management and ensure that all security software is properly controlled and audited.

7.5.8 Resource Control

Resource control activities include organization leadership, budget preparation, contract management, membership on the EMAP Management Board, and provide the primary external representative for EMAP IM. This is the responsibility of the Technical Director/Technical Coordinator in conjunction with each Resource and Coordinating Group information manager. This activity is continuous with yearly budget preparations and defenses. Key activities for the early stages of the project are staffing and identifying computer systems requirements.

7.5.9 Administration

The IM Coordinator is responsible for the oversight of administrative services. While most of the staff associated with EMAP IM reside within other EPA and non-EPA organizations, the IM Coordinator is responsible for maintaining contractual agreements for project work and is involved in hardware and software purchases and support decisions. The User Interaction and Planning function will identify training requirements and develop a training program and schedule to meet the required user skill levels. This involvement began with the POC and extends through full EMAP implementation. Deliverables will be in the form of training requirement documents and implementation schedules.

7.5.10 User Interaction

As products are created by the EMAP IM process, the User Interaction and Planning function will continue to work with the Resource and Coordinating Group users to validate, refine, and update user requirements. Through ongoing JAD sessions and direct user contact, future needs will be assessed and potential pilots will be identified. Pilots that are to be considered include prototypes on data distribution tracking, data control and access, support tracking and reporting, change requests, QA procedures, and archiving alternatives. For each prototype that is conducted, a JAD or joint meeting will be held to define design requirements.

Close coordination will be maintained with other applicable organizational functions such as the System Architecture group so that revisions can be made to the architecture. For each prototype, the User Interaction and Planning group will create data and process models as well as user interfaces for demonstrations. They also will identify requirements for standards, QA, and configuration management for all architecture components.

User interaction findings will be considered in each iteration of the strategic and tactical planning processes. Through this evolutionary system development process, user needs are continually refreshed and advanced technologies assessed on a continuing basis.

7.6 Implementation Schedules

As discussed in Section 5.2.3 the EMAP IM process development cycle embodies all phases of the SDLC. The evolutionary approach used in EMAP IM allows for the overlapping of these phases. The following schedules depict where the phases are combined and when these phases are to be accomplished.

7.6.1 EMAP IM Implementation Schedule – POC

The Implementation Schedule for the POC process is shown at Figure 7.6.1.1. The focus of this process is two Resource Groups, Forests and Estuaries, and is projected to be completed by FY 1995.

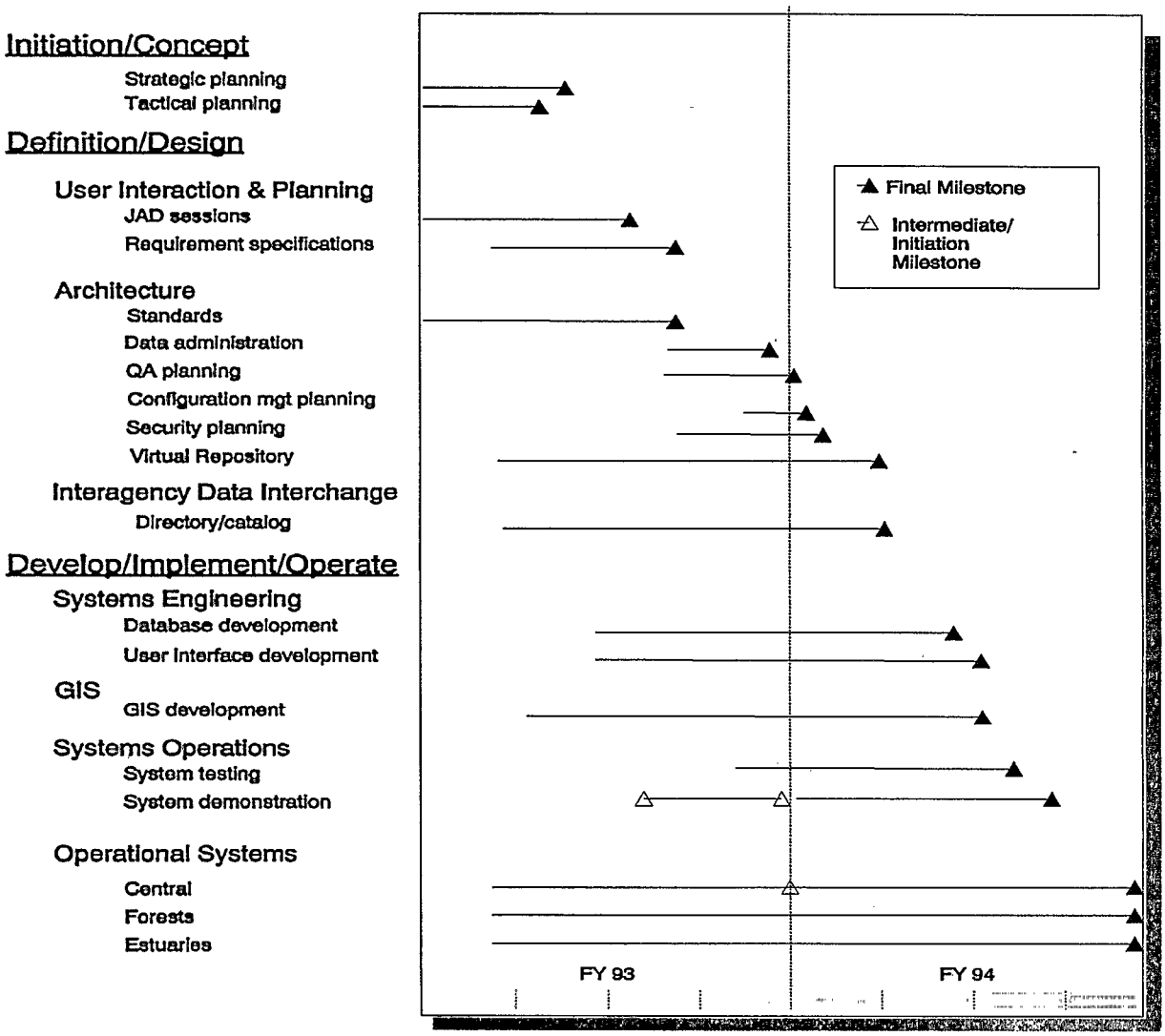

Figure 7.6.1.1. EMAP IM implementation schedule – POC.

7.6.2 EMAP IM Implementation Schedule – Technology Transfer

The Implementation Schedule for the Technology Transfer process is shown at Figure 7.5.2.1. This process is intended to include all Resource Groups and selected Coordinating Groups, involving increased program activity and an increased number of deliverables. This process is projected to be completed by FY 1996.

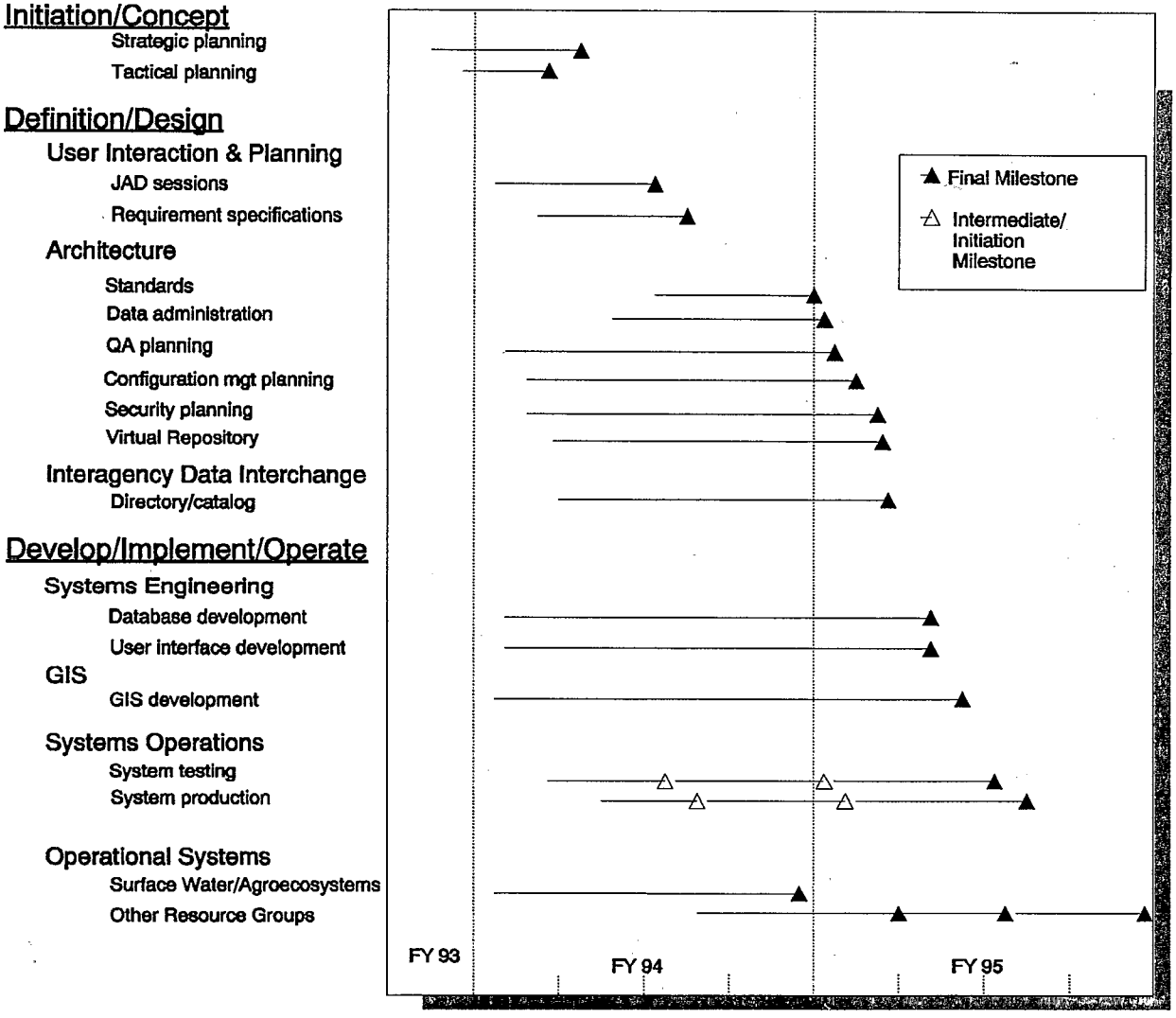

Figure 7.6.2.1. EMAP implementation schedule – Technology Transfer.

7.6.3 EMAP IM Implementation Schedule – Enterprise Implementation

The Implementation Schedule for the Enterprise Implementation process is shown at Figure 7.6.3.1. This process involves all Resource and Coordinating Groups and expands in scope to include the Integration and Assessment component, EMAP National Program within EPA, and the EMAP Interagency Federation. This process is projected to be completed by FY 1998.

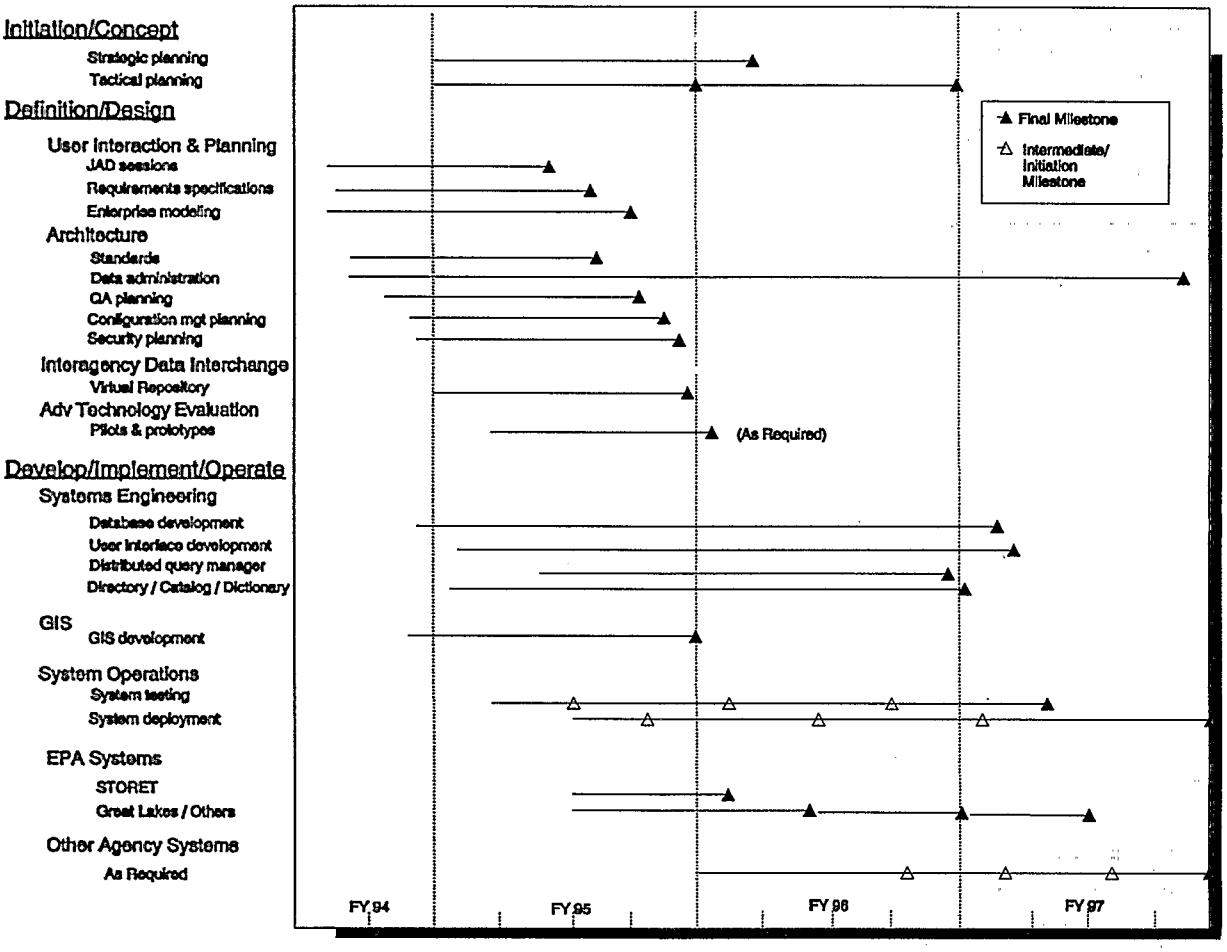

Figure 7.6.3.1. EMAP IM implementation schedule - Enterprise.

7.7 Capability by Year: User's Perspective

The sections below represent the type of capability planned for each year between 1993 and 1998. Each of these major capabilities will be the result of implementing (portions of) one or more of the IM systems architecture components.

7.7.1 EMAP IM Capability, 1993

The EMAP IM POC was developed for initial field test in 1993. This demonstration pilot enables two Resource Groups to collect, manage, and share data in a limited functionality environment. The POC is being used:

- To create operational systems for Forests and Estuaries;
- To develop a central node composed of corporate and warehouse databases;
- As the basis for technology transfer to other Resource and Coordinating Groups;
- As a demonstration tool for EMAP IM systems staff to convey the vision of EMAP to management and Cooperative Partners; and
- For improving and evaluating user interface and GIS software.

7.7.2 EMAP IM Capability, 1994

Technology Transfer will begin with the remaining Resource Groups. This process will integrate the system foundation from the POC into operational systems. The 1994 system will be capable of:

- Accepting data from field recording equipment;
- Maintaining an initial EMAP-wide Virtual Repository;
- Performing basic analysis using commercial survey analysis software;
- Offering an initial user interface based on the Proof-of-Concept;
- Offering initial tools for generating information products;
- Performing internal data exchange adhering to data standards; and
- Accessing EMAP IM data via a GIS.

7.7.3 EMAP IM Capability, 1995

Technology Transfer will be complete and the Enterprise Implementation process will begin. EMAP IM systems will be re-engineered to be compatible with other EPA programs, and integration planning will commence. EMAP IM will be capable of:

- Initial integration of improved data acquisition tools;
- Improved data verification;

- Offering a fully operational EMAP IM Virtual Repository;
- Implementing survey analysis methods developed for EMAP and providing interfaces to commercial survey analysis software;
- Offering an enhanced user interface featuring visualization;
- Access and limited reporting across ecological resources;
- Accommodating internal data transformation methods; and
- Providing a fully functional GIS.

7.7.4 EMAP IM Capability, 1996

The EMAP IM Enterprise process will focus on integrating the information management systems of the EMAP National Program within EPA with other agencies. It will be capable of:

- Full use of enhanced data acquisition tools;
- A Virtual Repository linking data and metadata;
- Enhanced analysis and aggregation methods;
- Offering an enhanced user interface featuring multimedia;
- Offering a common user interface across many platforms; and
- Providing for initial external data exchange.

7.7.5 EMAP IM Capability, 1997

During 1997, EMAP IM systems should be operated in the field with an increased number of Cooperative Partners. EMAP will have commenced assessments of IM systems in other agency and international organizations in order to create an Interagency Federation of environmental monitoring and assessment systems.

7.8 Long-Term System Operation, Evaluation & Enhancement

The scope of this Plan is limited to the period extending from the present through FY97. Beyond 1997, EMAP IM systems will continue to operate, and it is likely that system enhancements will continue to occur. Although beyond the scope of this Plan, some discussion of system operation and enhancement beyond 1997 is appropriate. At this time, it is expected that continuous system evaluation and enhancement will occur in the post-1997 time frame. Certainly, technology will evolve and improve during that time period, and infusing appropriate technology into EMAP IM systems will continue to be a priority of development. User needs will continue to evolve, and it is expected that in the future, data will be exchanged between EMAP IM systems and other programs to a greater degree.

Appendix

A

IRM COORDINATION

A.1 EMAP IM Systems within Federal IM and EPA

EMAP IM systems must be coordinated with the Federal and Environmental Protection Agency (EPA) Information Resources Management (IRM) programs. This participation will be discussed in Section A.1.1, Federal Roles and Responsibilities, and Section A.1.2, EMAP IM within EPA. Major Federal IRM guidelines that pertain to EMAP IM include Federal Information Processing Standards (FIPS), and information management guidelines from the General Accounting Office (GAO) and Office of Management and Budget (OMB).

A.1.1 Federal Roles and Responsibilities

The GSA has defined roles and responsibilities of Federal IRM functions as stated in the *Senior Federal IRM Manager*, a sample of which is illustrated in Table A.1.1.1.

IRM Functions		
Information Management	**Leadership Role In IRM**	**Promulgating Overall IRM Policies and Guidance**
Evaluate and improve the accuracy, completeness, and reliability of data and records contained within Federal information systems. Raise awareness of the importance of IM within the agency. Provide and implement the tools to help agency programs identify and manage information needs. These tools include standards, locator systems, and inventories. Provide leadership, serve as a catalyst, develop programs, define information requirements and develop common definitions. Increase the emphasis on the management of information.	Serve as agency advocate for promulgating and implementing the concept of IRM. Raise the awareness of IRM as being integral to mission needs. Emphasize to all agency senior-level staff that information management is part of their job. Serve as Agency representative to central oversight agencies such as OMB, GSA, Congress, GAO, etc. Coordinate with and participate in IRM councils, committees, and task forces that deal with multi-agency issues. Provide advice and council to key agency officials on IRM issues. Provide advice on agency procurements.	Develop and promulgate policies, procedures, and guidelines to ensure that IM resources are in support of the mission and managed effectively throughout the agency. Implement Federal laws and regulations. Implement an agency-wide IRM program that emphasizes innovative methods for managing information as a resource that must be planned, budgeted, and controlled just like any other resource. • Develop policies to ensure economical management of resources • Implement requirements for carrying out IM activities • Ensure compliance with departmental/agency standards for all IRM disciplines

Table A.1.1.1. Federal Roles and Responsibilities

A.1.2 EMAP IM within EPA

The relationship and responsibility of EMAP IM within the EPA illustrated in Figure A.1.2.1. Within the EPA, the Senior IRM Manager of the Office of Administration and Resources Management provides support and direction to the EMAP Management Board concerning IRM issues. The Senior IRM Manager works with the EMAP Management Board to coordinate budgets and to approve major acquisitions, strategic plans, systems design, and system documents.

The Senior IRM Manager works with the IM Coordinator to provide IRM leadership and to develop and promulgate IRM policies and guidelines. The Senior IRM Manager provides oversight, IRM reviews, and technical consultation and partici-

pates in discussions concerning new technologies, IM risks, and other relevant issues.

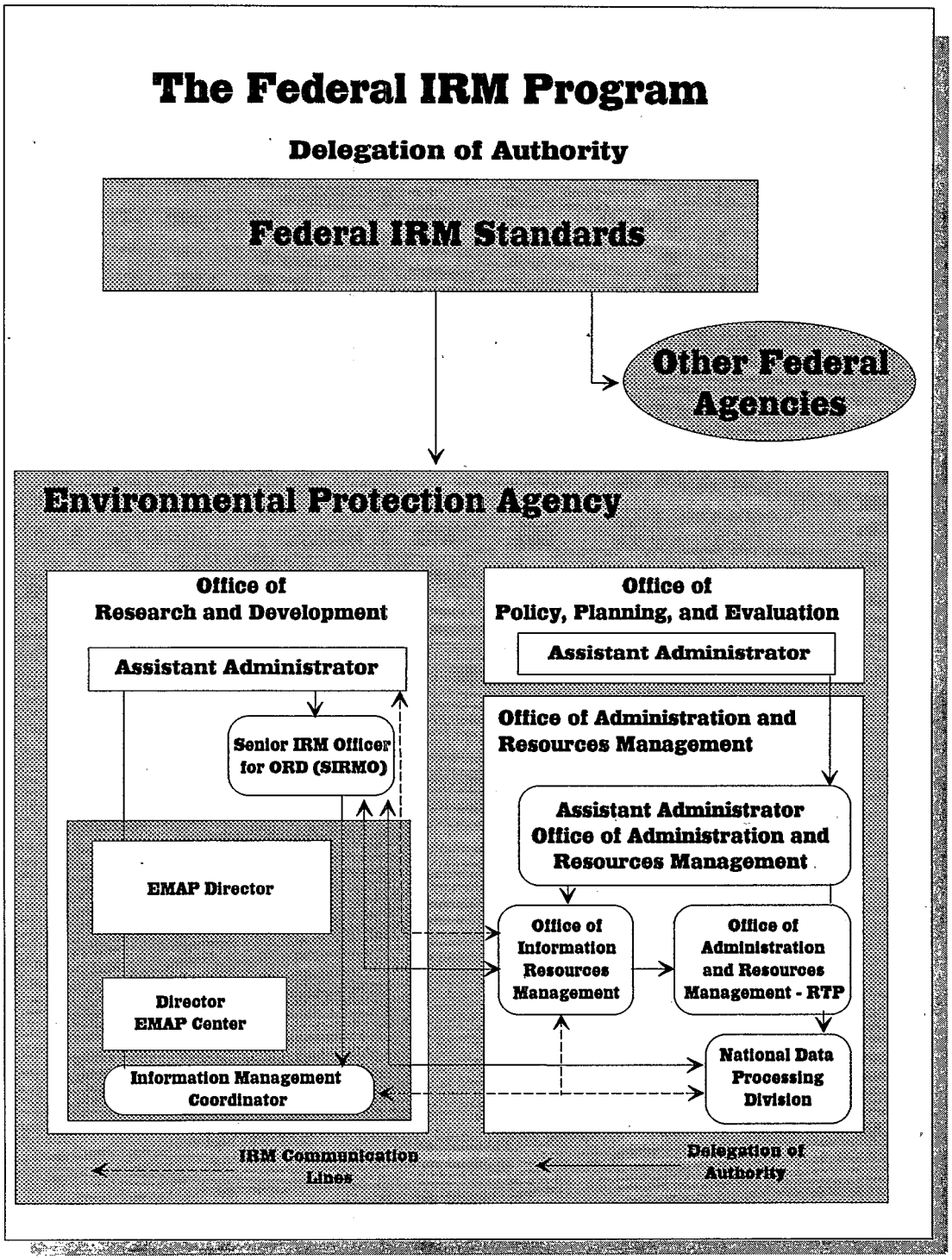

Figure A.1.2.1. Federal IRM/EPA/EMAP relationship.

The Senior IRM Manager may delegate authority to a Senior IRM Officer (SIRMO) to work with the organization and with the EMAP Resource and Coordinating Groups to perform IRM functions. The SIRMO may delegate IRM tasks to the IM Coordinator and/or other appropriate EMAP managers. In this document, the term "Senior IRM Manager" may refer to one or more delegated IRM officials.

Information Management/Data Administration (IM/DA), as shown in Figure A.1.2.2, will help ensure that agency standards are met whenever possible while still achieving EMAP objectives. IM/DA will provide EMAP with guidelines and emerging standards for data administration. The EMAP architecture and engineering functions will participate with the IM/DA group in all appropriate standards development groups such as National Institute of Standards and Technology (NIST), Open Systems Foundation (OSF), American National Standards Institute, Inc. (ANSI), and the International Standards Organization (ISO). IM/DA will routinely provide EMAP a list of all appropriate standards development activities. These collaborative efforts will provide critical information for many other projects in the EPA. Since few accepted standards exist for the management of scientific information, it will be necessary to work with other agencies such as the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), and the U.S. Department of Energy (DOE), where significant funds are being devoted to this type of development.

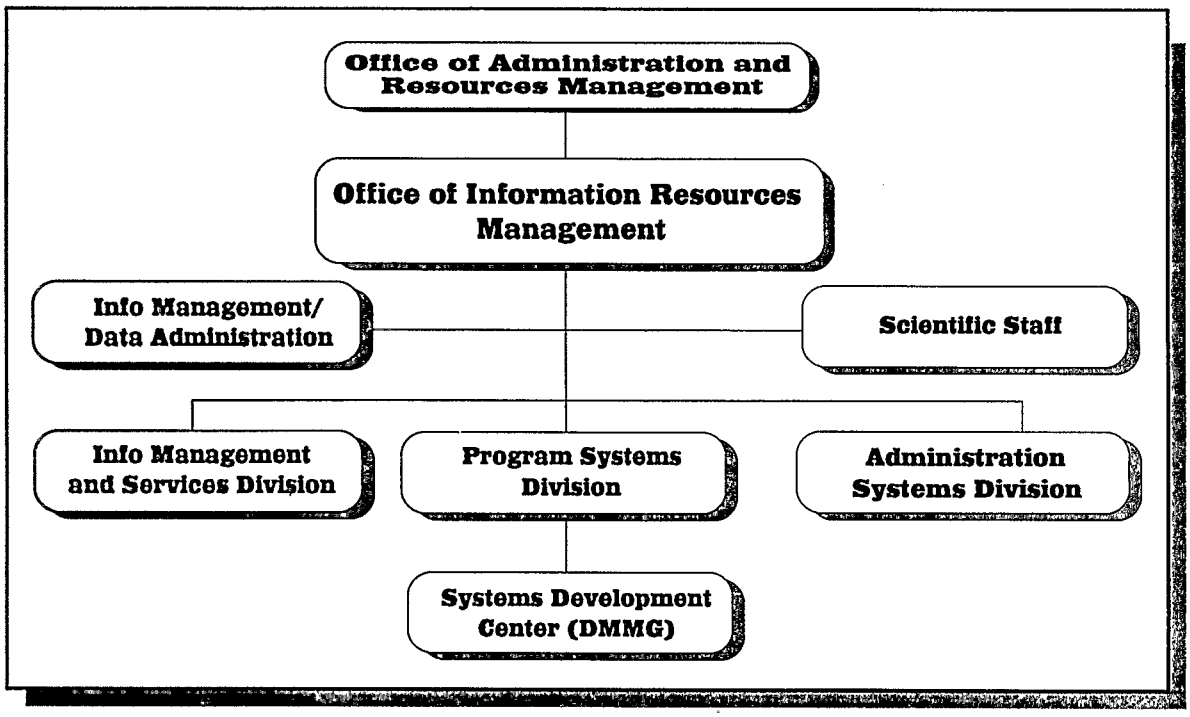

Figure A.1.2.2. OIRM organization.

EMAP IM systems development will assist the IM/DA group in setting reasonable and productive standards.

The System Design and Development work group within the Information Management Services Division is responsible for setting agency methods and system development strategies, and updating the System Design and Development Guidelines (SDDG) to current EPA methods for systems development. This is especially true for scientific information systems. EMAP will contribute to the update of the SDDG by attending work sessions, providing feedback, and making practical suggestions. The SDDG update will return with guidelines and tools to enable a complete, documented, and productive approach to EMAP IM systems development. Until such an accepted SDDG is developed, the EMAP IM systems development effort will proceed with the approach outlined in this Plan.

The **System Development Center** (SDC) will support the EMAP IM systems engineering development effort as EMAP becomes a National Program within EPA. The SDC, through the MOSES contract, will provide implementation resources that can develop systems to meet EMAP IM systems architecture specifications and acceptance test plans. EMAP will provide input and guidance to the SDC to ensure that the most efficient methods are used and that system components meet EMAP specifications.

In the **Office of Administration and Resources Management** (OARM) two groups within the National Data Processing Division will have regular interaction with EMAP IM systems development (see Figure A.1.2.3). These are the Architectural Management and Planning Branch (AMPB) and the Telecommunications Branch (TB). This interaction will focus on the physical aspects of EMAP IM systems, specifically hardware, commercial software, and networks. The OARM provides the operational support for all the EPA field laboratories and central computer centers. In addition, EMAP will rely on the OARM to provide budget support and actual financial support for all EPA shared computational, commercial software, and telecommunications that are needed.

EMAP will work with the AMPB to design and develop EMAP IM systems physical architecture in compliance with agency standards. If this hampers EMAP from achieving its goals then EMAP will work to enhance the operational environment with AMPB. AMPB will assist EMAP with architecture planning and procurement support.

EMAP will work with the Telecommunications Branch (TB) to specify network and communication requirements for EMAP IM systems. The TB must ensure that adequate capacity exists and network plans are in compliance with the agency standards. EMAP will provide information to TB to enable them to support the EMAP telecommunication needs. It is not anticipated that EMAP will need funda-

mentally different network communications than that currently provided or under consideration by TB.

All EMAP sites have network connections (EPA Information Technology Architecture). This is currently composed of a national SNA T-1 circuits network connecting EPA headquarters with its regional office facilities. Also EPA provides X.25 network connections among the regions' local computing resources. These resources will be used, in part, for the initial data analysis. In addition, many of the States have provided network extensions for their specific needs. These networks are also connected to external networks such as Internet. EMAP network requirements must be provided to the TB which will implement within budget limitations the required network infrastructure. Initial studies have indicated that major network improvements will not be necessary in the near term. As EMAP begins to integrate related information, including images, network requirements will change significantly.

In summary, EMAP will participate and cooperate with EPA OIRM groups to pilot development of agency methods, standards, and QA metrics; assist with development of an OIRM/EPA infrastructure that includes tools, encyclopedias, repositories, hardware, networks, etc.; lead in the use of an iterative design process (evolutionary development); explore the use of work products as an approach to project management; and integrate OIRM goals of achieving economies of scale in its procurements through considered of use of common hardware, software, and training and through standardization.

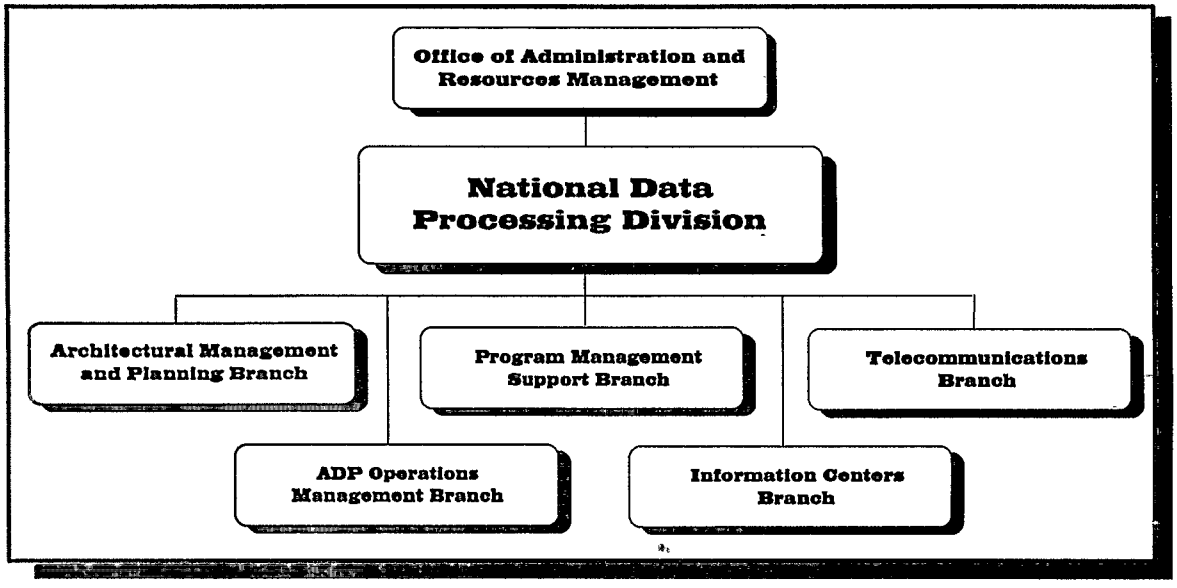

Figure A.1.2.3. NDPD organization.

Appendix

B

ORD ORGANIZATION

B.1 EMAP Relationship to Rest of ORD

The scope and complexity of EMAP requires extensive coordination among various organizations within EPA's Office of Research and Development (ORD). As shown in Figure B.1.1, the EMAP Program Director reports to the Director of the Office of Modeling, Monitoring Systems and Quality Assurance within ORD. Several Laboratories and Assessment Groups are in ORD and many interact with EMAP. The Resource Group Technical Directors in EMAP are organizationally responsible to the Directors of their respective Laboratory but are functionally responsible on EMAP-related matters to the EMAP Director. The Coordinating Groups in EMAP are directly responsible to the EMAP Director through the Chief of Integration and Assessments and the EMAP Center Director.

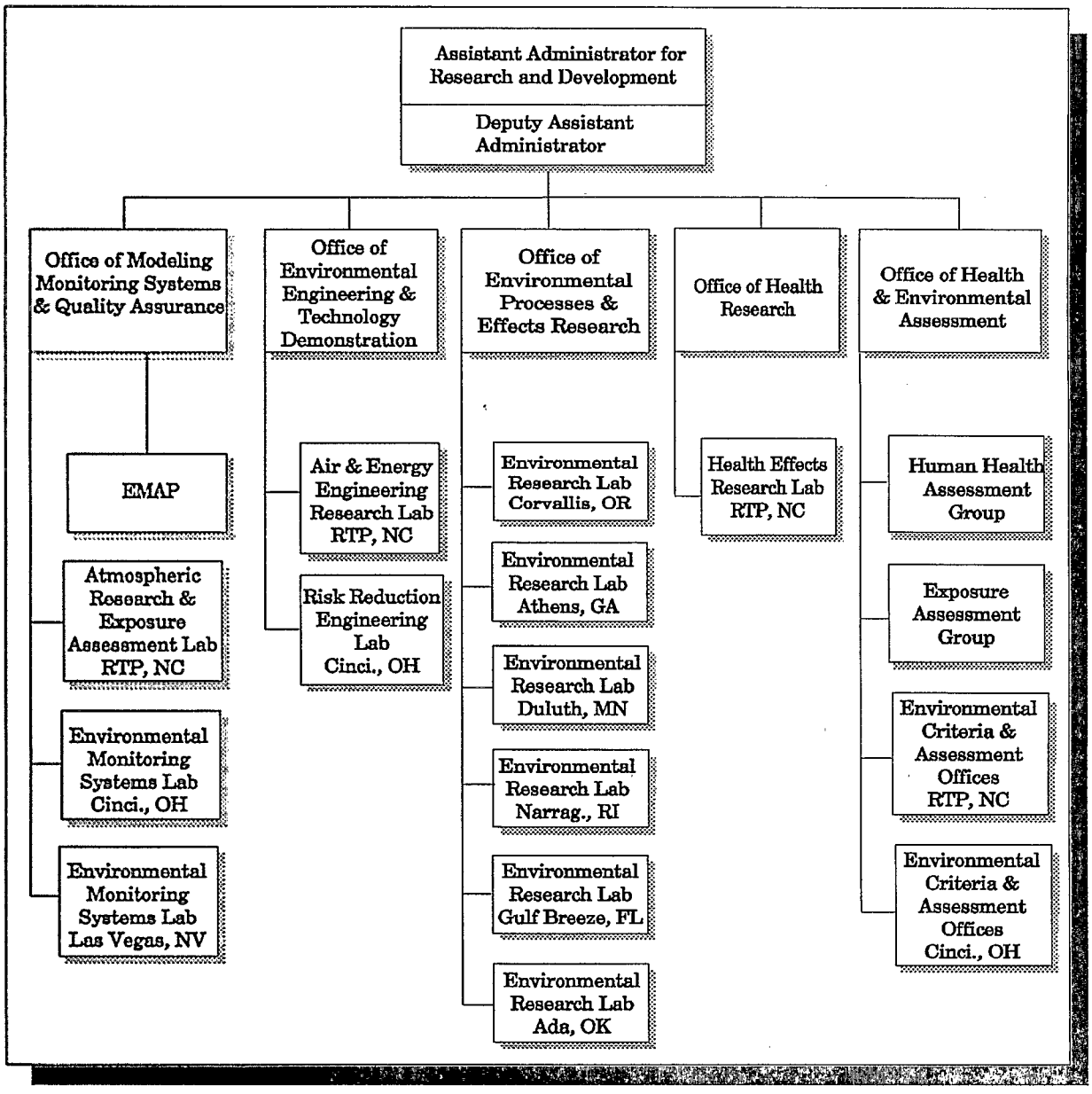

Figure B.1.1. Office of Research and Development organizational chart.

GLOSSARY

accuracy	The degree to which a calculation, a measurement, or set of measurements agree with a true value or an accepted reference value.[59]
aggregation	The process of collecting, deriving or summarizing.
agroecosystem	A dynamic association of crops, pastures, livestock, other flora and fauna, atmosphere, soils and water.[59]
ancillary data	Data collected from studies within EMAP but not used directly in the computation of an indicator.[59]
annual statistical summary	A document that presents a brief and comprehensive report of EMAP data collected on a single EMAP resource for a specific year.
arid ecosystem	Terrestrial systems characterized by a climate regime where the potential evapotranspiration exceeds precipitation, annual precipitation is not less than 5 cm and not more than 60 cm, and daily and seasonal temperatures range from -40°C to 50°C. The vegetation is dominated by woody perennials, succulents, and drought resistant trees.[59]
assessment	Interpretation and evaluation of EMAP results for the purpose of answering policy-relevant questions about ecological resources, including determination of the fraction of the population that meets a socially defined value, and association among indicators of ecological condition and stressors.[59]

attribute	Any property, quality, or characteristic of an entity.
back plane	The architecture infrastructure layer that allows different software tools to be used or "plugged in" within a common architecture.
cardinality rules	Specify the number of times or occurrence that a data entity can participate in a relationship with another entity.
catalog	Set of detailed documentation about data sets.
Census TIGER	Spatial information format used for Census Bureau data.
characterization	Determination of the attributes of resource units, populations, or sampling units.[59]
client/server	Use of distributed "client" computer systems linked to a central "server" in order to share common software tools/applications and capability.
completeness	The amount of valid data obtained compared to the planned amount.[59]
computer science metadata	Information describing the management of data, such as where it is stored, what format it is stored in, and what version it is.
conceptual	Abstract or generalized.
condition	The distribution of scores describing resource attributes without respect to any societal value or desired use, that is, a state of being.[59]
condition indicator	A characteristic of the environment that provides quantitative estimates of the state of ecological resources and is conceptually tied to a value.[59]
Cooperative Partners	Those Federal agencies that participate with EPA in EMAP. These include U.S. Department of Agriculture, U.S. Fish and Wildlife Service, U.S. Bureau of Land Management, National Oceanic and Atmospheric Administration, and others.
Coordinating Group	Group of scientific and administrative personnel headed by a technical coordinator and charged with addressing specific crosscutting integrative issues in EMAP, such as Landscape Characterization, Design and Statistics, Indicator Development, Information Management, Assessment

	and Reporting, Logistics, Methods and Quality Assurance.
cross-cutting group	In EMAP's Integration and Assessment section, one group of scientific and administrative personnel headed by a technical coordinator and charged with addressing specific cross-program, integrative issues in EMAP, such as Landscape Characterization, Design and Statistics, Indicator Development, Information Management, Assessment and Reporting, Logistics, Methods, and Quality Assurance.[59]
database	Information located in a relational database system.
data set	A logically meaningful grouping or collection of similar or related data.[60]
data warehouse	A database used for decision support.[47]
dictionary	Set of descriptions, formats, and other basic information about elements in a database.
directory	Set of summarized documentation about data sets.
distributed information system	Physically distributed set of systems for managing information.
domain	The areal extent of a region; the region occupied by a resource.[59]
ecological indicator	See condition indicator
ecoregion	Regions of relative homogeneity in ecological systems or in relationships between organisms and their environments.[59]
ecosystem	The interacting system of a biological community and its non-living environmental surroundings.[59]
EEI 1	Mission Needs Statement
Enterprise	The business entity consisting of operations, management, procurement, administration, sales, etc., and the functions such as information management which support it.
Enterprise architecture	Conceptual representation of an entire business.
Enterprise Implementation Process	The third phase of the stratified approach to EMAP IM development and implementation.

	Incorporates other EPA and non-EPA programs into the systems development process.
entity	A "thing" of importance.
entity relationship diagram	Graphical representation of data objects and their relationships.
environment	The sum of all external conditions affecting the life, development, and survival of an organism.[59]
environmental assessment	An environmental analysis prepared pursuant to the National Environmental Policy Act to determine whether a Federal action should significantly affect the environment and thus require a more detailed environmental impact statement.[59]
estuary	Regions of interaction between rivers and nearshore ocean waters, where tidal action and river flow mix fresh and salt water.[59]
extensible system	Automated/manual system that can be extended without being re-implemented.
forest	Land with at least 10% of its surface area stocked by trees of any size or formerly having had such trees as cover and not currently built-up or developed for agricultural use.[59]
geographic information system	A collection of computer hardware, software, and geographic data designed to capture, store, update, manipulate, analyze, and display geographically referenced data.[59]
Great Lakes	In EMAP, the resource that encompasses the five Great Lakes, wetlands contiguous to the lakes, and the connecting channels.[59]
heterogeneous	Consisting of dissimilar or diverse constituents.
indicator	In EMAP, characteristics of the environment, both abiotic and biotic, that can provide quantitative information on ecological resources.[59]
indicator development	The process through which an indicator is identified, tested, and implemented.[59]
integration	The formation, coordination, or blending of units or components into a functioning or unified whole. In EMAP, integration refers to a coordinated approach to environmental monitoring, research, and assessment. Integration in EMAP also refers to the technical processes involved in

	normalizing and combining data for interpretation and assessment.[59]
integration engineering	System implementation and integration.
Internet	International Communications Network
inventory	Contents of the Data Set Directory.
landscape	The set of traits, patterns, and structure of a specific geographic area, including its biological composition, its physical environment, and its anthropogenic patterns.[59]
landscape characterization	Documentation of the traits and patterns of the essential elements of the landscape, including attributes of the physical environment, biological composition, and anthropogenic patterns. In EMAP, landscape characterization emphasizes the process of describing land use or land cover, but also includes gathering data on attributes such as elevation, demographics, soils, physiographic regions, etc.[59]
landscape ecology	The study of distribution patterns of communities and ecosystems, the ecological processes that affect those patterns, and changes in pattern and process over time.[59]
measurement	A quantifiable attribute that is tied to an indicator.[59]
metadata	Descriptive or qualifying data that describes primary data elements.
metadata classification	The type of metadata. For example: metadata related to the science at hand versus metadata about the processes used to aggregate/derive the data.
model	Mathematical or physical representation of data or a system that accounts for all or some of its known properties.[59]
Model Manager	A tool that manages models, such as function hierarchies, network diagrams, database schemas, data flow diagrams, process models, and entity-relationship diagrams.
monitoring	In EMAP, the periodic collection of data that is used to determine the condition of ecological resources.[59]

open systems	An automated system that can be accessed and extended openly.
pilot	Development of a subset of complete-system functionality that will evolve, at least in part, into a total system. Pilots will be distributed for testing and evaluation.
precision	The degree to which replicate measurements of the same attribute agree or are exact.[59]
proof-of-concept	A prototype developed to test and prove ideas.
Proof-of-Concept (POC) Process	The first phase of a stratified approach for EMAP IM development and implementation. The POC process identifies and applies standards, policies, tools, and procedures to the Forests and Estuaries Resource Groups and limited Coordinating Groups.
prototype	Developed system of functionality for the purpose of evaluating new technology. Also used to interactively gather requirements.
quality assessment	The evaluation of environmental data to determine if they meet the quality criteria required for a specific application.[59]
quality assurance (QA)	An integrated system of activities involving planning, quality control, quality assessment, reporting and quality improvement to ensure that a product or service meets defined standards of quality with a stated level of confidence.[59]
quality control (QC)	The overall system of technical activities whose purpose is to measure and control the quality of a product or service so that it meets the needs of users.[59]
rapid application development	The process used by EMAP IM to expedite computer systems functionality through prototyping to the users.
rapid prototyping	Quickly creating a pseudo-functional system that embodies user-defined capabilities.
raw data	Data that is not derived or aggregated. A basic atomic piece of information.
Resource Group	In EMAP, one of eight ecological entities or ecosystem types that shares certain basic characteristics. These are: Estuaries, Great

	Lakes, Lakes and Streams, Wetlands, Forests, Arid Ecosystems, Agroecosystems, and Landscape Ecology.
region	Any explicitly defined geographic area. In the EMAP objectives, region refers to the ten standard Federal regions.[59]
relation	The concept of function, correlation, or association between or among attributes, which may be qualitative as well as quantitative.[59]
scientific metadata	Information describing scientific data, such as how, when, and where it was collected.
SNA T1 circuit	System Network Architecture communication circuit.
software back plane	Common software environment that other software packages can be connected to.
software pilot	Creation of (a set of) software that will be fielded and used. Will be evolved/enhanced into fully operational system.
STORET	Information management system developed by the Office of Water for standardizing and monitoring water data.
stressor	Any physical, chemical, or biological entity that can induce an adverse response.[59]
surface waters	The inland surface waters consisting of all the nation's lakes (other than the Great Lakes), rivers, and streams.[59]
system architecture	The overall logical and physical definition of a system.
system design	Specifications for implementation of a system.
system development life cycle	Chronological process used for fielding a completed system.
system prototype	See rapid prototyping.
T1, T2, T3	High speed data communications trunk.
technical coordinator (TC)	The individual responsible for directing the activities of an individual cross-cutting group.[59]
technical director (TD)	The individual responsible for directing the activities of an individual resource group.[59]
technology transfer	The process of sharing technology validated during an initial test phase to other operations.

Technology Transfer Process	The second phase of a stratified approach for EMAP IM development and implementation. This process extends the activities validated during the POC process to all Resource Groups and an increasing number of Coordinating Groups.
trends	The changes in the distribution of scores for condition indicators over multiple time periods.[59]
validation	The process of substantiating specified performance criteria.[59]
verification	The process of ensuring correctness.
Virtual Repository	A collection of metadata describing EMAP data and information systems, plus a set of tools that makes metadata available to users and system developers.
Wetlands	An area that is saturated by surface or ground water with vegetation adapted for life under soil conditions such as swamps, bogs, fens, marshes, and estuaries.[59]
X.25	Packet switching communications protocol
Zachman Framework	The "checklist" used by EMAP IM for implementing the system development life cycle.

LIST OF ACRONYMS

ADP	Automated Data Processing
AMPB	Architectural Management and Planning Branch
ANSI	American National Standards Institute, Inc.
ASCII	American Standard Code for Information Interchange
ATE	Advanced Technology Evaluation
BLM	Bureau of Land Management
CAS	Chemical Abstracts Service
CASE	Computer Aided Software Engineering
CD ROM	Compact Disk Read Only Memory
CES	Center for Environmental Statistics
DAAC	Distributed Active Archive Center
DBMS	Database Management System
DCE	Distributed Computing Environment
DMMG	Development and Maintenance Methodology Group
DOE	Department of Energy
EMAP	Environmental Monitoring and Assessment Program
EMAP IM	Environmental Monitoring and Assessment Program Information Management
EMSL-LV	Environmental Monitoring Systems Laboratory, Las Vegas

EOS	Earth Observing System
EOS/DIS	Earth Observing System/Distributed Information System
EPA	Environmental Protection Agency
ERD	Entity Relationship Diagram
Federal IM	Federal Information Management
FICCDC	Federal Interagency Coordinating committee for Digital Cartography
FIPS	Federal Information Processing Software
FTE	Full-time Equivalent
FURPS	Functionality, Usability, Reliability, Performance System
GAO	General Accounting Office
GIS	Geographic Information Systems
GNMP	Government Network Management Protocol
GOSIP	Government Open Systems Interface Protocol
GSA	Government Services Administration
GRD	Geographic Reference Data
GUI	Graphical User Interface
HW	Hardware
IAG	Interagency Agreement
ICB	Information Centers Branch
IDI	Information Dimensions, Inc.
IM/DA	Information Management/Data Administration
IRM	Information Resource Management
ISA	Information Systems Architecture
ISO	International Standardization Organization
JAD	Joint Application Design
MOSES	Mission Oriented System Engineering

MOU	Memorandum of Understanding
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NDPD	National Data Processing Division
NETCDF	Network Common Data Format
NIST	National Institute of Standards and Technology
NOAA	National Oceanographic and Atmospheric Administration
NRC	National Research Council
NSF	National Science Foundation
OARM	Office of Administration and Resources Management
OIRM	Office of Information Resource Management
OMB	Office of Management and Budget
ORD	Office of Research and Development
OSF	Open Systems Foundation
OSWER	Office of Solid Waste and Emergency Response
PMB	Program Management Branch
POC	Proof-of-Concept
POSIX	Portable Operating System Standard
QA	Quality Assurance
QC	Quality Control
RAD	Rapid Application Development
RDA	Remote Data Access
RDBMS	Relational Database Management System
SAB	Science Advisory Board

SAS	Statistical Analysis System
SCS	Soil Conservation Service
SDC	System Development Center
SGML	Standard Generalized Markup Language
SDLC	System Development Life Cycle
SDDG	System Design and Development Guideline
SDTS	Spatial Data Transfer Standard
SIRMO	Senior Information Resources Management Officer
SNMP	Simple Network Management Protocol
SQL	Structured Query Language
SW	Software
TB	Telecommunications Branch
TCP/IP	Transmission Control Protocol / Internet Protocol
TVA	Tennessee Valley Authority
UI	User Interface
USDA	United States Department of Agriculture
USFS	United States Forest Service
USGCRP	United States Global Climate Research Program
WAIS	Wide Area Information Servers

REFERENCES

[1] Environmental Monitoring and Assessment Program: Program Guide, June 1993.

[2] *The National Information Infrastructure: Agenda for Action*, September 15, 1993.

[3] Zachman, J.A., *A Framework for Information Systems Architecture*, IBM Systems Journal, Vol. 26, No. 3, 1987.

[4] *IBM: A Management System for the Information Business*, Vol. 1 - Management Overview, GE 20-0662-1, 1984.

[5] Office of Solid Waste and Emergency Response (OSWER) *System Life Cycle Management Guidance*. Dir. No. 9028.00.

[6] Martin, James. *Rapid Application Development*, Macmillan Publishing Company, New York, 1991.

[7] EMAP Information Management Task Group, System Life Cycle Management Studies Manual, (draft), U.S. Environmental Protection Agency, 1993.

[8] Science Advisory Board, *Future Risk: Research Strategies of the 1990s*, CAB-EC-88-040, Washington, DC, U.S. Environmental Protection Agency, 1988.

[9] EEI-1 Mission Needs Statement EMAP, July 1991.

[10] Ecological Indicators, EPA/600/3-90/060, September 1990. Edited by C.T. Hunsaker and D.E. Carpenter

[11] EMAP Landscape Characterization Information Requirements Summary, August 1990.

[12] EMAP IM Strategic Plan Peer Review Summary, Washington, DC, March 1993.

[13] French, J.C., et. al., Scientific Data Management, NSF Workshop Report on Scientific Data Management, Computer Science Report No. TR-90-21, Department of Computer Science, University of Virginia, August 1990.

[14] Rosen, J.S., J. Beaulieu, M. Hughes, H. Buffum, J. Copeland, R. Valente, J. Paul, F. Holland, S. Schimmel, C. Strobel, K. Summers, K.J. Scott, J. Parker. 1990. *Data Base Management System for Coastal Demonstration Project.* EPA Office of Research and Development. Internal Report. EPA/600/x-90/207.

[15] FY91-93 ADP Requirements EMAP, April 1991.

[16] Bromley, Allan, *Data Management for Global Change Research Policy Statements*, U.S. Global Change Research Program, National Science Foundation, DOE/EP-0001P, July 1991.

[17] Adamus, P.R. and R.J. O'Connor. 1991. *Quality Assurance Project Plan: 1991-92 Northeast Bird Studies (EMAP/Biodiversity).* U.S. Environmental Protection Agency, Corvallis, OR.

[18] Peck, D.V. (ed.). 1991. *Environmental Monitoring and Assessment Program: Integrated Quality Assurance Project Plan for the Surface Waters Resource Group.* EPA/600/X-91/080. U.S. Environmental Protection Agency, Las Vegas, NV.

[19] Peck, D.V. (ed.). (In preparation.) *Environmental Monitoring and Assessment Program: Integrated Quality Assurance Project Plan for the Surface Waters Resource Group. Revision 2.* U.S. Environmental Protection Agency, Las Vegas, NV.

[20] Valente, R. M., C. J. Strobel, J. E. Pollard, K. M. Peres, T. C. Chiangand J. R. Rosen. 1990. *Quality Assurance Project Plan for EMAP - Near-Coastal: 1990 Demonstration Project.* U. S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

[21] Heitmuller, P.T., R. Valente, and J.K. Summers. 1991. *Near-Coastal Louisianian Province 1991 Monitoring Demonstration - Quality Assurance Project Plan.* U.S. Environmental Protection Agency, Office of Research and Development. EPA/600/1-91/xxx.

[22] Valente, R. M. and J. R. Schoenherr. 1991. *EMAP - Near-Coastal 1991 Virginian Province Quality Assurance Project Plan.* U. S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI.

[23] Heitmuller, P.T. and J.K. Summers. 1992. *Quality Assurance Plan for 1992 Sampling Season-EMAP/Estuaries – Louisianian Province.*

[24] Thomas, J.J., et al. 1979. *Analysis of Large Data Sets on a Minicomputer.* Proceedings of Computer Science and Statistics: 12th Annual Symposium on the Interface. University of Waterloo, Ontario, Canada. pp. 442-446.

[25] Thomas, J.J., Burnett, R.A., and Lewis, J.R. 1981. *Data Editing on Large Data Sets.* Proceedings of the Computer Science and Statistics: the 13th Symposium on the Interface. pp. 252-258.

[26] Thomas, J.J., 1982. *The Impact of Large Data Sets on Interaction Style for Data Manipulation Languages.* Proceedings of the First LBL Workshop on Statistical Database Management. pp. 157-159.

[27] Thomas, J.J., 1982. *A User Interaction Model for Manipulation of Large Data Sets.* Proceedings of the Computer Science and Statistics: 14th Symposium on the Interface. Troy, NY.

[28] Connell, John, *Structured Rapid Prototyping,* Yourdon Press, 1989.

[29] Wiley, John, *Application Prototyping: A Requirement Definition Strategy for the 80s,* Bernard Boar, 1984.

[30] Gause, D.C. and G.M. Weinberg, *Exploring Requirements: Quality Before Design,* Dorset House Publishing, 1989.

[31] DeMarco, Tom, *Structured Analysis and System Specification,* New York: Yourdon Press, 1978.

[32] Flavin, M., *Fundamental Concepts of Information Modeling,* New York: Yourdon Press, 1981.

[33] Yourdon, Edward, and Larry L. Constantine, *Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design,* New York: Yourdon Press, 1979.

[34] Lantz, Kenneth S., *The Prototyping Methodology*, Englewood Cliffs, NJ: Prentice Hall, 1986.

[35] Cernosek, G., E. Monterio, and E.W. Pribyl, *An Entity-Relationship Approach to Software Requirements Analysis for Object-Based Development*, Houston, TX: McDonnell Douglas Astronautics, 1987.

[36] Coad, P. and E. Yourdon, *Object-Oriented Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

[37] Davis, C., S. Jajodia, P. Ng, and R. Yeh, eds., *Entity-Relationship Approach to Software Engineering*, Amsterdam, The Netherlands: Elsevier Science, 1983.

[38] Freeman, P., *Software Systems Principles*, Chicago, IL: Science Research Associates, 1975.

[39] Glass, R., *Modern Programming Practices: A Report from Industry*, Englewood Cliffs, NJ: Prentice Hall, 1982.

[40] Jackson, M., *System Development*, Englewood Cliffs, NJ: Prentice-Hall, 1983.

[41] Bertino, E., *Distributed Database Design Using the Entity-Relationship Model, in Entity-Relationship Approach to Software Engineering*, ed. C. Davis et al, Amsterdam, The Netherlands: Elsevier Science, 1983.

[42] Braithwaite, K., *An Implementation of a Data Dictionary to Support Databases Designed Using the Entity-Relationship (E-R) Approach*, In: Entity-Relationship Approach to Software Engineering, ed. C. Davis, et al, Amsterdam, The Netherlands: Elsevier Science, 1983.

[43] Wood, Jane, and Denise Silver. *Joint Application Design.* John Wiley and Sons, New York, 1989.

[44] Boehm, B.W., *A Spiral Model of Software Development and Enhancement.* In Software Engineering Project Management, IEEE, Inc., 1987.

[45] *Managed Evolutionary Guidebook.* Office of the Assistant Commissioner for Information Systems, U.S. Patent and Trademark Office, Arlington, VA, June 1993.

[46] Sowa & Zachman, *Extending and Formalizing the Framework for Information Systems Architecture*, IBM Systems Journal, Vol. 31, No. 3, 1992.

[47] Gallagher, Bob, "Data Warehouse in Store for Downsizing Firms; Distribution Method Causes IS Headaches." PC Week, Vol. 10, No. 1, Jan. 11, 1993.

[48] Strebel, D.E. and J.B. Frithsen. *Handling Supporting Information for EMAP External Data Sets.* Versar, Inc., Columbia, MD. December 31, 1991.

[49] EMAP IM POC Standards Manual (draft), U.S. Environmental Protection Agency, July 1993.

[50] *Directory Interchange Format Manual. Version 4.0.* NASA, National Space Science Data Center, Greenbelt, MD. December 1991.

[51] Spewak, Steven H. and Steven C. Hill, *Enterprise Architecture Planning*, QED Publishing Group, Boston, 1993.

[52] Forum on Data Management for Global Change - November 2-4, 1988. Baltimore, MD. Committee on Earth Sciences – Interagency Working Group on Data Management for Global Change (IWGDMGC).

[53] *The FIFE Information System*, Strebel, et al, Trans. GARS 1990.

[54] *Information Systems Support for Interdisciplinary Field Experiments*, Strebel, et al, S Proc AMS 1991.

[55] *Earth Science Information Systems*, Mason, et al, Proc IST Conference on ESSIS.

[56] *Our Changing Planet: The FY 1993 U.S. Global Change Research Program, A Supplement to the U.S. President's Fiscal Year 1993 Budget.*

[57] *Solving the Global Change Puzzle - U.S. Strategy for Managing Data and Information* (NRC).

[58] EMAP Information Management Task Group. *Entity Relationship Diagrams for EMAP Resource Groups* (draft). U. S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory - Las Vegas, NV. July 27, 1992.

[59] EMAP. *Environmental Monitoring and Assessment Program: Master Glossary.* EPA/620/R-93/013, Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, Environmnetal Monitoring and Assessment Program. October, 1993.

[60] Strebel, D.E. and J.B. Frithsen. *Draft Metadata Guidance for the EMAP Proof of Concept.* Versar, Inc., Columbia, MD. February 5, 1993.

U.S. GOVERNMENT PRINTING OFFICE: 1994 - 550-001/80410