



EMAP Information Management Plan: 1998–2001



**Environmental Monitoring
and Assessment Program**

EMAP

Information Management Plan: 1998–2001

by

Stephen Hale¹, Jeffrey Rosen², Dillon Scott²,
John Paul¹, and Melissa Hughes³

¹ Atlantic Ecology Division, U.S. Environmental Protection Agency,
27 Tarzwell Drive, Narragansett, RI 02882

² Technology Planning and Management Corporation, Mill Wharf
Plaza, Suite 208, Scituate, MA 02066

³ OAO Corporation, U.S. Environmental Protection Agency,
27 Tarzwell Drive, Narragansett, RI 02882

Contract Number 68-W5-0034

National Health and Environmental Effects Research Laboratory
Office of Research and Development
U. S. Environmental Protection Agency

Notice

The U.S. Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under Contract Number 68-W5-0034 to Technology Planning and Management Corporation. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is contribution number NHEERL-NAR-2018 of the Atlantic Ecology Division.

Abstract

This Information Management Plan describes how the Environmental Monitoring and Assessment Program (EMAP) manages data and information to support EMAP research and policy objectives. The plan includes descriptions of EMAP data, data users, the processes and technology with which users can access EMAP data, and the infrastructure that supports these activities.

Key words: environmental monitoring, environmental assessment, information management, information systems, systems architecture, systems engineering, database management system, USEPA-EMAP, information resources management, geographic information system.

Preface

The Environmental Monitoring and Assessment Program (EMAP) is a research program whose goal is to:

“Monitor the condition of the Nation’s ecological resources to evaluate the cumulative success of current policies and programs and to identify emerging problems before they become widespread or irreversible. (U.S. EPA 1997b)”

EMAP researchers analyze, assess, and report on large volumes of spatial and temporal ecological data and information. The Information Management Plan outlined below provides a vision, scope, approach, and resource requirements for EMAP information management to support the EMAP activities. Information management goals outlined in this plan follow the principles stated in the draft EMAP Research Plan (U.S. EPA 1997b) and will evolve as the EMAP Research Plan develops.

The Information Management Plan has gone through three stages: the development of the initial EMAP Information Management Strategic Plan: 1993–1997 (Shepanek 1994); a 1996 updating of this Plan that reflects the 1995 change in mission and scope of the EMAP program (U.S. EPA 1996a); and this 1998 update. This version revises and strengthens the previous plan by incorporating the results of requirements analysis meetings with EMAP Working Groups and completing Essential Elements of Information (EEI) documentation requirements. These revisions are in accordance with the comments made by the inter-agency EMAP Data Management Review Team in Baltimore, May 1996 (U.S. EPA 1996e).

The structure and content of this document provides a common departure point for receiving constructive feedback about system design and implementation plans, and for developing consensus on the mission and direction of EMAP Information Management. It is a living document that will provide a primer for learning about EMAP Information Management and a baseline description for the system as it evolves in response to new program requirements and improved system technologies.

Contents

Notice	ii
Abstract	ii
Preface	iii
List of Acronyms	xiii
Acknowledgments	xviii
EMAP Information Management Plan Executive Summary	xix
Section 1 Introduction and Approach	1
1.1 Purpose and Scope of the EMAP Information Management Plan	1
1.2 Intended Audience for the EMAP Information Management Plan	2
1.3 Development of the EMAP Information Management Plan	3
1.4 EMAP Background	4
1.5 Early EMAP	4
1.6 Current EMAP	5
1.7 EMAP Information Management	7
1.7.1 EMAP–IM Mission	8
1.7.2 EMAP–IM Approach	9
1.7.3 Role of EMAP–IM (AED)	13
1.8 EMAP Data Policy Statement	13
1.9 Conclusions	18
Section 2 EMAP Data	19
2.1 Introduction	20
2.2 EMAP 1990–1995 Data	20
2.3 Current EMAP 1996–2001 Data	22
2.3.1 ORD Regional-Scale Assessments Program—Mid-Atlantic Integrated Assessment and Western Pilot	26
2.3.2 Intensive/Index Sites	32
2.3.3 Landscape Ecology	40
2.3.4 Regional EMAP	44

Contents

2.3.5	Ecological Indicator Development	47
2.4	Conclusions	52
Section 3 Information Management Needs and Requirements		53
3.1	Introduction	55
3.2	User Needs	55
3.2.1	General Users	56
3.2.2	Primary Users	57
3.3	Recommended Guidelines for EMAP Data Sources	58
3.3.1	Types of EMAP Data Sources	59
3.3.2	Recommended Guidelines for Data Management and Delivery	59
3.4	EMAP–IM Functional Requirements	65
3.4.1	Track EMAP and Non-EMAP Data Relevant to EMAP Research	65
3.4.2	Facilitate Rapid, Ad Hoc Data Exchange Among EMAP Researchers	67
3.4.3	Provide Standards, Tools, and Support to Users and Data Collectors	67
3.4.4	Maintain and Update EMAP–IM System (Components and Network Connections)	69
3.4.5	Deliver EMAP Data and Information Managed by EMAP–IM (AED)	69
3.5	System Requirements	69
3.5.1	Overall System Design	70
3.5.2	System Components	70
3.5.3	System Configuration—Software, Hardware, Network, and Online Resources	73
3.6	Conclusions	75
Section 4 Technical Design		76
4.1	Purpose	77
4.2	Background of EMAP–IM System Development	78
4.2.1	Early EMAP Information Management System (1990–1995)	78
4.2.2	Current EMAP Information Management System (1996–)	79
4.3	System Concept and Overview of Technical Structure	79
4.4	System Components	80
4.4.1	EMAP Data Directory	81
4.4.2	EMAP Data Catalog	87
4.4.3	EMAP Public Web Site	91

Contents

4.4.4	EMAP Internal Web Site	93
4.4.5	EMAP Summary Data Sets	93
4.5	System Configuration	94
4.6	EMAP Archival Plan	97
4.7	System Evaluation	97
4.7.1	Data Accessibility	97
4.7.2	Flexibility of Design to Adapt to Future Technological and Program Changes	98
4.7.3	User Satisfaction	98
4.7.4	Benefits and Costs	99
4.7.5	Risks and Contingencies	99
4.8	Need for System Enhancement	99
4.9	Conclusions	100
Section 5 Project Management and Coordination		101
5.1	Introduction	102
5.2	EMAP Program Management Structure	103
5.2.1	EMAP Management Structure	103
5.2.2	EMAP Program Management Related to Non-EPA Research Partners	104
5.3	EMAP–IM Project Management Structure	104
5.4	Information Management in Working Groups	106
5.5	Relationship Between EMAP–IM and Information Management Authorities	107
5.5.1	Coordination within EPA	107
5.5.2	Coordination with Other Federal Groups	110
5.6	EMAP–IM Project Management Challenges	113
5.6.1	Time Frames for EMAP Data Availability	113
5.6.2	Minimum Requirements for Data Delivery	113
5.6.3	Budgeting for Information Management	113
5.6.4	Ensuring that the EMAP–IM System is Fulfilling User Needs	115
5.6.5	Availability of Resource Group Data	115
5.6.6	Capture of Data Sets that have No Long-Term Stewards	115
5.6.7	Developing Effective Relationships with Other Data Repositories	116
5.6.8	Data Exchange Among EMAP Researchers	116
5.6.9	Assistance to Working Groups for Version Tracking and Documentation	117
5.7	Conclusions	118

Contents

Section 6 Implementation Plan	119
6.1 Introduction	121
6.2 System Components—Maintenance and Enhancement	121
6.2.1 Data Directory	121
6.2.2 EMAP Data Catalog	124
6.2.3 EMAP World Wide Web Site	126
6.2.4 EMAP Internal Web Site	130
6.3 Early EMAP (1990–1995) Data	131
6.3.1 Management and Coordination	131
6.3.2 Tasks	131
6.3.3 Responsibilities	132
6.3.4 Resources Needed	132
6.4 Current EMAP (1996–) Data	132
6.4.1 Management and Coordination	132
6.4.2 Tasks	133
6.4.3 Responsibilities	134
6.4.4 Resources Needed	134
6.5 Data Management	135
6.5.1 Management and Coordination	135
6.5.2 Tasks	135
6.5.3 Responsibilities	136
6.5.4 Resources Needed	136
6.6 GIS Spatial Data and Analyses	136
6.6.1 Management and Coordination	136
6.6.2 Tasks	136
6.6.3 Responsibilities	136
6.6.4 Resources Needed	136
6.7 Mid–Atlantic Integrated Assessment	136
6.7.1 Management and Coordination	137
6.7.2 Tasks	137
6.7.3 Responsibilities	138
6.7.4 Resources Needed	138
6.8 Western Pilot Regional Assessment	138
6.8.1 Management and Coordination	138
6.8.2 Tasks	138

Contents

6.8.3	Responsibilities	139
6.8.4	Resources Needed	139
6.9	EMAP–IM System Administration and Coordination	139
6.9.1	Tasks	139
6.10	Overall Resource Requirements for EMAP–IM (AED)	140
6.11	Conclusions	141
6.12	Tentative Schedule	141
6.12.1	FY1999 Tasks	141
6.12.2	FY2000 Tasks	143
6.12.3	FY2001 Tasks	145
References		146
Glossary		154

Contents

(Appendices under separate cover)

Appendix A Essential Elements of Information Requirements Report	1
A.1 Purpose	1
A.2 EEI-1 Mission Needs Analysis	2
A.3 EEI-2 and EEI-3 Preliminary Design and Options Analysis	4
Appendix B Data Management Needs and Practices of EMAP Working Groups	7
B.1 Purpose	8
B.2 EMAP and Working Group Mission and Goals	8
B.3 Requirements Analysis Overview	8
B.4 ORD Regional-Scale Assessments—Mid-Atlantic Integrated Assessment Pilot	10
B.5 Intensive/Index Sites	20
B.6 Landscape Ecology	29
B.7 Regional EMAP (R-EMAP)	46
B.8 Ecological Indicator Development	73
B.9 Committee on the Environment and Natural Resources	80
Appendix C Inventory of EMAP Data	91
C.1 Purpose	91
C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data ...	92
C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data	109
C.4 Types, Volumes, and Status of Other Data	116
Appendix D Preliminary Design and Options Document	118
D.1 Purpose	118
D.2 Option—Enhancement to EMAP Oracle Database to Handle Complex Data Types .	119
D.3 Conclusions and Next Steps	123
Appendix E Responses to “Environmental Monitoring and Assessment Program: Data Management Review Team Report”	125
E.1 Background	125
E.2 Review Team Comments and EMAP–IM (AED) Responses	125
E.3 Review Team Members	143

Contents

Appendix F Overview of EMAP Information Management Policies, Guidelines, and Standards 144

F.1 Introduction 144

F.2 Data Sharing 145

F.3 EMAP Public Web Site 145

F.4 EMAP Data Directory 145

F.5 EMAP Data Catalog 145

F.6 Further Information 146

Appendix G EPA IRM Vision Elements 149

Appendix H Configuration of the Computing Infrastructure of the Atlantic Ecology Division and National EPA 151

Appendix I EMAP Archival Plan 155

I.1 Introduction 155

I.2 Requirements for EMAP Data Storage and Usability 156

I.3 Types of Data Comprising EMAP 156

I.4 Current Digital Data Backup/Archival Scheme 156

I.5 Long-Term Goals for Digital Data Archives 157

I.6 EMAP Digital Archive Tape Validity Testing 158

I.7 Migrating EMAP Data to New Hardware and Software 158

I.8 EMAP Archival Tracking System 158

Appendix J Organization of ORD Offices and Laboratories 159

Appendix K Contributors to the Development of the EMAP IM System 161

K.1 EMAP Information Management Working Group 161

K.2 Contributors to EMAP Information Management, 1989-1998 161

Appendix L Partial Bibliography for EMAP IM Program 164

List of Tables and Figures

Tables

Table 1-1.	Early and Current EMAP Research Groups	4
Table 1-2.	How EMAP–IM Objectives Support EMAP Program Objectives	10
Table 1-3.	Existing EMAP Information Management System Components	12
Table 2-1.	Repositories of 1990–1995 Raw and Summary Data Sets	22
Table 2-2.	Working Group Partners Responsible for Data Sets	23
Table 2-3.	Location of Raw and Summary Data for Working Groups	25
Table 2-4.	MAIA–Estuaries Data Management	28
Table 2-5.	MAIA–Surface Waters Data Management	30
Table 2-6.	MAIA–Landscape Ecology Data Management	32
Table 2-7.	EMAP Intensive Sites—DISPro—UV-B Monitoring	35
Table 2-8.	Intensive Sites—Coastal Intensive Sites Monitoring Data Management	40
Table 2-9.	Landscape Ecology Data Management	42
Table 2-10.	R-EMAP Projects by Region	44
Table 2-11.	R-EMAP Overall Program Data Management	46
Table 2-12.	Ecological Indicator Development Data Management	49
Table 3-1.	User Categories	56
Table 4-1.	Types of EMAP Data and the Corresponding EMAP System Components Used to Manage Them	81
Table 4-2.	List of All Attributes in the Directory Database	85
Table 4-3.	EMAP Data Catalog fields	89
Table 4-4.	U.S. EPA (AED) Software Environment Supporting EMAP	96
Table 4-5.	U.S. EPA (RTP) Software Environment Supporting EMAP	97

List of Tables and Figures

Figures

Figure 4-1.	Components of the distributed EMAP–IM system	79
Figure 4-2.	Flow of data and metadata from data sources to EMAP World Wide Web Site	81
Figure 4-3.	Flow of Data Directory information from data sources to EMAP World Wide Web Site	83
Figure 4-4.	Entity Relationship Diagram for EMAP Data Directory	85
Figure 5-1.	Structure of EMAP Organization within ORD	106
Figure 5-2.	Relationship of EMAP to EPA’s OIRM and ORD’s ORMA	109

List of Acronyms

ADP	Automated Data Processing
AED	Atlantic Ecology Division
AIRMoN	Atmospheric Integrated Research Monitoring Network
AIRS	Aerometric Information Retrieval System database (EPA)
ANSI	American National Standards Institute
ARS	U.S.D.A. Agricultural Research Service
ASCII	American Standard Code for Information Interchange
AVHRR	Advanced Very High Resolution Radar (satellite imagery)
BLM	Bureau of Land Management
CAS	Chemical Abstracts Service
CASE	Computer Aided Software Engineering
CASTNet	Clean Air Status and Trends Network
CBAT	Community-Based Assessment Team
CBP	Chesapeake Bay Program
C-CAP	Coastal Change Analysis Program (NOAA)
CCRS	Canada Centre for Remote Sensing
CD-ROM	Compact Disk Read-Only Memory
CENR	Committee on the Environment and Natural Resources (White House)
CES	Center for Environmental Statistics
CIESIN	Consortium for International Earth Science Information Network
CISNet	EPA-NOAA Coastal Intensive Sites Network
CSC	NOAA Coastal Services Center
CSV	Comma-separated values (in ASCII format)
DAAC	Distributed Active Archive Center
DBMS	Database Management System
DCE	Distributed Computing Environment
DEM	Digital elevation models (USGS elevation data)

Acronyms

DIF	NASA Directory Interchange Format
DISPro	NPS Demonstration Intensive Sites Project
DMMG	Development and Maintenance Methodology Group
DOE/ORNL	Department of Energy Oak Ridge National Laboratory
DRG	Digital Raster Graph (USGS digital quadrangle format)
EDC	USGS EROS Data Center satellite imagery processing, archiving and distribution data center (Sioux Falls, SD)
EEI	Essential Elements of Information (EPA Directive 2182)
EIMS	Environmental Information Management System (EPA database)
EMAP	Environmental Monitoring and Assessment Program
EMAP-IM	EMAP Information Management
EMAP-IM (AED)	EMAP Information Management Staff, Atlantic Ecology Division
ENVI	The Environment for Visualizing Images image processing software
EROS	Earth Resources Observation System (partnership of USGS, NASA, NOAA, ICSU, US AID, EPA, Department of Defense & Intelligence Community, UNEP/GRID)
EOS	Earth Observing System
EOS/DIS	Earth Observing System/Distributed Information System
EPA	Environmental Protection Agency
ERD	Entity Relationship Diagram
ETSD	Enterprise Technology Services Division (EPA RTP)
FGDC	Federal Geographic Data Committee
FHM	Forest Health Monitoring
FICCDC	Federal Interagency Coordinating committee for Digital Cartography
FIPS	Federal Information Processing Software
FTE	Full-time Equivalent
FTP	File Transfer Protocol
FURPS	Functionality, Usability, Reliability, Performance System
GAO	U.S. General Accounting Office
GAP	Gap Analysis Program (USFWS)
GB	Gigabyte (1,073,741,824 bytes)
GCDIS	Global Change Data and Information System
GCMD	Global Change Master Directory
GCRIO	Global Change Research Information Office

Acronyms

GCRP	Global Change Research Program
GED	Gulf Ecology Division (ORD)
GIS	Geographic Information System
GSA	U.S. Government Services Administration
GRD	MAIA Geographic Reference Database
GUI	Graphical User Interface
HTML	HyperText Markup Language
HUC	USGS Hydrologic Unit Code
IAG	Interagency Agreement
IM	Information Management
IMPROVE	Interagency Monitoring of Protected Visual Environments Program (NPS)
IMWG	Information Management Working Group (EMAP)
IRM	Information Resource Management (EPA)
ISO	International Standards Organization
LAN	Local Area Network
LANDSAT-MSS	Landsat Multi-Spectral Scanner
LANDSAT-TM	Landsat Thematic Mapper
LTER	Long-Term Ecological Research (NSF)
MAIA	Mid-Atlantic Integrated Assessment (ORD Regional Assessments)
MB	Megabyte (1,048,576 bytes)
MDN	Mercury Deposition Network
MED	Mid-Continent Ecology Division (ORD)
MOU	Memorandum of Understanding
MRLC	Multi-Resolution Land Characteristics Consortium (partners: EMAP, NALC, GAP, NAWQA, C-CAP, EDC, RSCA, states)
NALC	EPA North American Land Characterization (partners: NASA, EDC, CCRS)
NADP/NTN	National Atmospheric Deposition Program/National Trends Network
NALC	North American Land Characterization
NAMS	National Air Monitoring System
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NAWQA	USGS National Water Quality Assessment Program
NBS	National Biological Survey (USGS)

Acronyms

NCEA	National Center for Environmental Assessment
NDDN	National Dry Deposition Network
NDVI	Normalized Difference Vegetation Index (changes in greenness)
NERL	National Exposure Research Laboratory (ORD)
NHEERL	National Health and Environmental Effects Research Laboratory (ORD)
NISO	National Information Standards Organization
NIST	National Institute of Standards and Technology
NOAA	National Oceanic and Atmospheric Administration
NPDES	National Pollutant Discharge Elimination System
NPS	National Park Service
NRC	National Research Council
NRCS	Natural Resources Conservation Service (USDA)
NRI	Natural Resources Inventory (NRCS)
NSF	National Science Foundation
NSTC	National Science and Technology Council
NUVMC	National Ultra-Violet Monitoring Center (UGA)
NWIS II	USGS National Water Information System
OARM	Office of Administration and Resources Management (EPA)
OIRM	Office of Information Resource Management (EPA)
OMB	Office of Management and Budget
ORD	Office of Research and Development (EPA)
ORMA	Office of Resource Management and Administration (EPA ORD)
OSF	Open Systems Foundation
OSTP	Office of Science and Technology Policy (White House)
POC	Proof-of-Concept
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
QA/QC	Quality Assurance/Quality Control
QAPP	EPA Quality Assurance Project Plan
QC	Quality Control
RAMAS	Risk Assessment Management Software
RDBMS	Relational Database Management System
ReVA	Regional Vulnerability Assessment (NERL)
R-EMAP	Regional EMAP Working Group

Acronyms

RSCA	USFS Remote Sensing Applications Center
RTP	Research Triangle Park, North Carolina (office of EPA)
SAB	Science Advisory Board
SAS	Statistical Analysis System (software package)
SCS	Soil Conservation Service
SDLC	System Development Life Cycle
SDDG	System Design and Development Guideline
SDTS	Spatial Data Transfer Standard
SIMCorB	Science Information Management Coordination Board (in ORD)
SIRMO	Senior Information Resources Management Officer (in OIRM)
SOP	Standard Operating Procedure
SQL	Structured Query Language
STORET	EPA Storage and Retrieval of U.S. Waterways Parametric Data
TB	Terabyte (1,099,511,627,776 bytes)
TCP/IP	Transmission Control Protocol / Internet Protocol
TFODM	CENR Task Force on Observations and Data Management
TVA	U.S. DOE Tennessee Valley Authority
UGA	University of Georgia
USDA	United States Department of Agriculture
USFS	United States Forest Service
USFWS	United States Fish and Wildlife Service
USGCRP	United States Global Climate Research Program
USGS	United States Geological Survey
UV-A	ultraviolet A
UV-B	ultraviolet B
UVMN	Ultraviolet Monitoring Network
WAIS	Wide Area Information Servers
WAN	Wide Area Network
WED	Western Ecology Division (ORD)
WOUDC	World Ozone and UV Radiation Data Centre
WWW	World Wide Web

Acknowledgments

We gratefully acknowledge the many individuals who have contributed to EMAP information management since the program began in 1989. Many of the people involved are listed in Appendix K. The current Plan has benefitted from the advice of the EMAP Information Management Working Group (Appendix K.1) and the other EMAP Working Groups who supplied information about the data management needs and practices of their programs. The data management work has been supported by the EPA Information Technology contractors at the Atlantic Ecology Division and other labs. This plan has been substantially improved by review comments provided by the interagency EMAP Data Management Review Team organized by Joe Alexander (ORD Deputy Director for Research) in May 1996. Lastly, we thank Bob Shepanek and Jeff Frithsen of the ORD National Center for Environmental Assessment and Larry Rossner, Barbara Brown, and David Bender of the Atlantic Ecology Division for suggesting improvements to earlier drafts.

EMAP Information Management Plan Executive Summary

Environmental Monitoring and Assessment Program

The Environmental Monitoring and Assessment Program (EMAP) represents a long-term research commitment by the U.S. Environmental Protection Agency (EPA) to develop the tools needed to assess and document the status and trends of the Nation's ecological resources. The EMAP mission, as described in the EMAP Research Plan (U.S. EPA, 1997b) is to:

“Monitor the condition of the Nation's ecological resources to evaluate the cumulative success of current policies and programs and to identify emerging problems before they become widespread or irreversible.”

This EMAP Information Management Plan (EMAP-IM Plan) describes the current approach and implementation of data and information management for the 1996–2001 program. Existing capabilities and upcoming enhancements are described. A summary of the approach is given by Hale et al. (1998).

Background

During 1990–1995, EMAP was a national ecological monitoring program designed to assess the condition of the Nation's natural resources and contribute to decisions on environmental protection and management. Data collection and analysis were conducted by the EPA Office of Research and Development (ORD) researchers, contractors, and cooperators using EMAP sampling and analytical tools, and following EMAP standards and protocols. These researchers were organized into Resource Groups. A central information management group (Central EMAP-IM) was formed to support the national monitoring program by leading data management and maintenance of a centralized database in accordance with an EMAP strategic plan. Central EMAP-IM developed a set of EMAP-IM system components, including a Data Directory, Data Catalog, Oracle database, and a web site, and conducted information management research.

The EMAP program was redirected by ORD in 1995 and is no longer a national monitoring program with centralized data management conducted exclusively by EPA staff. Data management is decentralized among research projects led by a diverse set of cooperating EPA and non-EPA

researchers organized into Working Groups. Working Groups integrate both research and monitoring data from federal, state, local, and academic sources for ecological assessments, and are responsible for data management and maintenance. EMAP-IM's role is to coordinate access to the data and to provide data management and distribution guidance, standards, and assistance to the Working Groups as needed.

Another key feature of the current EMAP program is its participation in the federal Committee on Environment and Natural Resources (CENR). CENR promotes the integration of environmental monitoring data from many sources to support assessment of regional and national trends in environmental quality. The EMAP-IM system will evolve in accordance with standards adopted by CENR to maximize integration of EMAP data with the data of other participating agencies.

Current EMAP Information Management Approach and Structure

EMAP-IM is composed of three main groups:

- **EMAP-IM (AED)**—Consists of EPA employees and contractors at EPA's National Health and Environmental Effects Research Laboratory's (NHEERL) Atlantic Ecology Division (AED). EMAP-IM (AED) coordinates data access and maintains the EMAP-IM system;
- **EMAP Information Management Working Group (IMWG)**—Includes representatives from NHEERL, the National Exposure Research Laboratory (NERL: Research Triangle Park, and Las Vegas), the National Center for Environmental Assessment (NCEA), and ORD Headquarters. The IMWG, chaired by an EPA employee from AED, provides direction and priorities for EMAP-IM; and
- **Working Groups**—Includes researchers participating in projects who collect, maintain, document, distribute, and forward data or hyperlinks to EMAP-IM (AED) for distribution on the EMAP Public Web Site.

EMAP-IM relies on the Working Groups for primary data processing and management, including quality assurance and the preparation of documentation. Working Group researchers collect and manage EMAP data at their sites and follow a variety of standards and procedures for data processing. EMAP-IM (AED) concentrates its efforts in four main areas: 1) maintaining directories of data and documentation on the EMAP Public Web Site to ensure access to relevant data; 2) participating in development and adoption of data standards that facilitate integration of data into EMAP assessments and the CENR framework, and encouraging their use by researchers; 3) providing assistance and leadership to Working Group researchers for information management issues; and 4) technology transfer to regional programs like the Mid-Atlantic Integrated Assessment (MAIA).

EMAP Data 1996–2001

Since 1996, data collection, analysis, and distribution activities have increased in scope and complexity from those of the earlier program. EMAP data now span a wider array of disciplines, natural resource types, and methods of data collection and aggregation. EMAP data and information products now include monitoring data and tools for methodology and analysis, data aggregations created in support of integrated assessments, and documentation of these products. Data sets are held at many locations and processed by researchers using many different data management methods and standards. Data are also available in many different formats, including flat files, databases, maps, data sheets, and graphics. The EMAP–IM Plan summarizes the status of EMAP Working Group data that are being tracked and linked to EMAP users via the Internet.

Requirements

Requirements for the EMAP–IM system fall into four categories: user needs; recommended guidelines for data sources; functional requirements; and system requirements.

User needs are considered at two levels—primary and general. Primary users comprise the EMAP scientific community and data analysts. Primary users have clearly defined data needs for collecting, managing, documenting, and distributing EMAP data. Their use of the EMAP–IM system includes accessing and exchanging data with other researchers and locating relevant data sources for planning EMAP research. Planning for the EMAP–IM system is principally driven by the needs of primary users. General users include non-EMAP researchers, government managers, policy makers, and the general public. They use the EMAP–IM system to locate quality-assured EMAP information products and documentation.

Guidelines for data sources include recommendations for preparation and delivery of EMAP data by researchers, including: quality assurance/quality control (QA/QC); documentation; aggregation and integration; exchange among researchers; distribution to publicly accessible data repositories; and archival and storage.

Functional requirements include the EMAP–IM system components and technical expertise that EMAP-IM must provide to ensure the flow of EMAP Resource Group and Working Group data and metadata from the data sources to the end users. The core requirements of the system are maintenance of the Data Directory and a EMAP Public Web Site to track relevant data sources; facilitation of data exchange among researchers and delivery to end users; guidance and assistance to data sources with data standards, management, and tools; and maintenance of the EMAP–IM system.

System requirements are the physical characteristics of the EMAP–IM system that must be in place to fulfill the user needs and functional requirements outlined above. These requirements are being met by maintaining and enhancing the existing EMAP–IM components to track distributed data sets, improving the flow and delivery of data, increasing the accessibility of the EMAP Data Directory through the Internet, and ensuring interoperability with other environmental information systems such as the EPA Environmental Information Management System (EIMS).

Technical Design

EMAP–IM (AED) is the EMAP–IM system network coordinator, maintaining EMAP–IM system components (Data Directory, Data Catalog, EMAP Public Web Site, and Internal Web Site) that have been upgraded from the early EMAP program. The function of these components is:

- **EMAP Data Directory**—allows users to find data of interest by providing information about the location and accessibility of data sets. It consists of an Oracle database on the EMAP Public Web Site which can be accessed by web browsers.
- **EMAP Data Catalog**—provides an EMAP standard format for detailed metadata so that users can understand, correctly interpret, and use data. It consists of ASCII or HTML files on the EMAP Public Web Site.
- **EMAP Public Web Site**—provides the primary mechanism for linking users to EMAP data and information via the Internet. It consists of a set of linked web pages providing access to Resource Group and Working Group summaries, the EMAP Data Directory, EMAP Data Catalog files, EMAP–IM standards, EMAP publications, and hyperlinks to related web sites containing data and information of interest to EMAP users.
- **EMAP Internal Web Site**—a site on the EPA internal network, which is only accessible to users accessing the site from EPA computers. The site houses the maintenance version of the Data Directory, a Directory entry tool, a query interface for the Data Directory, and preliminary data, metadata, and documents that are being developed or reviewed for public release.

These components—in particular the Data Directory and EMAP Public Access Web Site—provide the foundation for EMAP to comply with EPA and CENR information distribution requirements.

Program Management

EMAP is organized as a core management group with a number of organizationally independent research partners that contribute funding and in-kind contributions to cooperative projects. The cooperative nature of the program means that EMAP is conducted in a matrix management environment in which control of data is distributed among organizations that perform the majority

of the work. EMAP-IM's goal is to track EMAP data and link it to EMAP users via the Internet, and the emphasis of the program will be on providing metadata and the tools for tracking the data. This task is complex because of the decentralized nature of the program among independent agencies. EMAP-IM has developed approaches that encourage standardization of data management and delivery, altruistic participation, and effective policies and operations to overcome these organizational deficiencies. EMAP-IM must encourage its EPA and non-EPA partners to document data and comply with emerging Federal environmental data and metadata standards. EMAP-IM must also coordinate with information management authorities such as the Office of Information Resource Management (OIRM), ORD Office of Resources Management and Administration (ORMA), the ORD Science Information Management Coordination Board (SIMCorB), and with other Federal information management authorities, including CENR.

EMAP-IM supports effective EMAP program management by:

- recommending information management standards and guidance in budgets and interagency agreements;
- taking the lead on preparation of Working Group and Research Group Data Directory entries and metadata;
- coordinating data and directory standards with other large-scale information management and dissemination efforts within EPA and other Federal agencies (e.g., CENR, Federal Geographic Data Committee [FGDC]); and
- encouraging EMAP data sources (research partners) to ensure long-term stewardship and distribution of data in publicly accessible systems with demonstrated longevity and success in managing, maintaining, and disseminating data (e.g., STORET).

Implementation

Implementation includes maintenance and enhancement of the EMAP-IM system components, creation of data and metadata references for the Data Directory and EMAP Public Access Web Site, guidance and assistance with EMAP standards, coordination with emerging Federal environmental information standards (e.g., FGDC, Global Change Research Program [GCRP]), and administration and coordination of information management tasks across the program. EMAP-IM (AED) leads the effort on all modifications and provides coordination with the IMWG, which reviews changes before they are implemented. Two of the core priorities for implementation are to complete documentation of EMAP data sets in the Data Directory and enhance the functionality of the EMAP Data Directory and the EMAP Public Access Web Site.

Planned activities include:

- **Data Directory:** enhance search capabilities and accessibility to the Internet, and comply with Z39.50 standard;
- **Data Catalog:** upgrade to database format to add searching capabilities;
- **EMAP Public Access Web Site:** enhance site to include more database capabilities, add EMAP Bibliographic Database, and improve delivery of spatial data;
- **Internal Web Site:** add capability to allow access to authorized research partners outside the EPA domain;
- **EMAP Data Delivery:** complete documentation for remaining Resource Groups and assist Working Groups with preparation and documentation of data;
- **Data Management:** provide support for development of databases such as the Aquatic Mortality Network and the EMAP-Estuaries Atlantic Coast database;
- **Internal EMAP-IM System Management:** upgrade EMAP tracking software, standards, and other internal tools;
- **Spatial Data Delivery:** expand capabilities to deliver spatial data on the EMAP Public Access Web Site;
- **Technology Transfer:** provide information management assistance to the MAIA and Western Pilot Study programs; and
- **EMAP Information Management Plan:** update to reflect changes to the program and needs for completing the information management mission.

Section 1

Introduction and Approach

- 1.1 Purpose and Scope of the EMAP Information Management Plan**
- 1.2 Intended Audience for the EMAP Information Management Plan**
- 1.3 Development of the EMAP Information Management Plan**
- 1.4 EMAP Background**
- 1.5 Early EMAP**
- 1.6 Current EMAP**
- 1.7 EMAP Information Management**
 - 1.7.1 EMAP–IM Mission**
 - 1.7.2 EMAP–IM Approach**
 - 1.7.3 Role of EMAP–IM (AED)**
- 1.8 EMAP Data Policy Statement**
- 1.9 Conclusions**

Information management is an integral part of the Environmental Monitoring and Assessment Program (EMAP) that provides a means for sharing and preserving data and information for all users well into the future. This section provides an introduction to the EMAP program and the priorities for information management that are presented in this Plan.

1.1 Purpose and Scope of the EMAP Information Management Plan

EMAP represents a long-term research commitment by EPA to develop the tools needed to assess and document the status and trends of the Nation’s ecological resources. The mission of the program is to:

“Monitor the condition of the Nation’s ecological resources to evaluate the cumulative success of current policies and programs and to identify emerging problems before they become widespread or irreversible.” (U.S. EPA 1997b).

The purpose of research and monitoring in the program is:

“...to develop the scientific understanding for translating environmental monitoring data from multiple spatial and temporal scales into assessments of ecological condition and forecasts of the future risks to the sustainability of our natural resources.” (U.S. EPA 1997b).

This EMAP Information Management (IM) Plan outlines the technical and project management approach that has been chosen to support this purpose and fulfill EMAP’s evolving information management needs. The Plan is intended to provide information for the implementation and maintenance of a system that serves the research, assessment, and data management needs of the current EMAP program and the Working Groups.

This Plan covers the implementation period from 1998–2001, and focuses primarily on the user needs and information systems technology that are relevant through 2001, including:

- how EMAP–IM supports EMAP program objectives;
- how EMAP data are being made available to users;
- the requirements and priorities of EMAP’s data users;
- guidelines for data originators;
- EMAP–IM functional and system requirements;
- project management issues for ensuring the effectiveness of the system; and
- an implementation plan describing activities for maintaining and enhancing the system under currently understood requirements over the next three years.

1.2 Intended Audience for the EMAP Information Management Plan

This document is written for program managers, senior information system managers, and scientists. The Plan will require regular updates as the scope and responsibilities, the user needs, and the technologies available to EMAP evolve. It is intended to help in the planning, management, implementation, and use of the EMAP–IM system. As such, it supports the following activities:

- providing U.S. Environmental Protection Agency (EPA) management with estimates of the level of effort and resources needed to implement effective information management in EMAP;

- establishing a common set of standards and expectations for planning research and data management for EMAP scientists and research partners in other agencies, universities, and industry;
- presenting and maintaining a road map for EPA information managers to use in directing system development and maintenance activities;
- providing other government agencies with information about EMAP's IM approach to facilitate data exchange and research cooperation on related environmental programs; and
- communicating information about the development, documentation, and implementation of the plan to the EPA Office of Research and Development (ORD) Senior Information Resources Management Officer (SIRMO).

1.3 Development of the EMAP Information Management Plan

This Information Management Plan has been prepared following EPA System Design and Development Guidance (U.S. EPA 1995c). Fulfillment of this guidance is documented in Appendix A (Essential Elements of Information Requirements Report). Requirements for documentation are summarized in Appendices B through F.

A 1996 Draft EMAP-IM Plan was reviewed by an interagency EMAP Information Management Review Team (U.S. EPA 1996e) and was updated to produce a draft revised version (U.S. EPA 1996a). The current version is the result of further revisions made to address the reviewers comments. It has been prepared by reviewing the draft Plan (U.S. EPA 1996a) and the Working Group research plans and by conducting a series of Requirements Analysis interviews with selected EMAP Working Groups.

Development of the IM Plan also included analysis of available and emerging standards and technologies. Standards for data management, documentation, and distribution were evaluated for their consistency with the Federal interagency Committee on the Environment and Natural Resources (CENR) and EPA existing requirements, and for their ability to improve data integration (see Section 4, Technical Design). Technologies were evaluated for their availability through existing EPA contracts and their ability to meet the requirements of the EMAP Working Groups and CENR.

The following sections present an overview of EMAP and the approach to information management that is being implemented to support the program.

1.4 EMAP Background

The early EMAP program ran from 1990–1995 as an intensive monitoring and data collection effort within ORD. The monitoring program was supported by research that included development of indicators, statistical methods, and field tools. In 1995, the program was redirected to a research program aimed at improving monitoring approaches including data integration from research partners in many agencies and from historical data. The components of the early and current program are outlined in Table 1-1.

Table 1-1. Early and Current EMAP Research Groups

EARLY EMAP 1990–1995 RESOURCE GROUPS	CURRENT EMAP 1996–2001 WORKING GROUPS
<ul style="list-style-type: none"> • Estuaries • Surface Waters • Wetlands • Forests • Rangelands • Great Lakes • Agroecosystems • Landscape Characterization 	<ul style="list-style-type: none"> • Regional-Scale Assessments (Mid-Atlantic Integrated Assessment Program, MAIA) • Index/Intensive Sites • Landscape Ecology • Regional EMAP (R-EMAP) • Ecological Indicator Development • EMAP Information Management Working Group (IMWG)¹

¹ The EMAP Information Management Working Group (IMWG) includes representatives of NHEERL (National Health and Ecological Effects Laboratory), NERL (National Exposure Research Laboratory), NCEA (National Center for Environmental Assessment), and the EMAP Working Groups. IMWG members represent EMAP on the CENR.

For more information about group activities and members, see the EMAP Public Web Site (EMAP 1998).

1.5 Early EMAP

Early EMAP was a national monitoring program designed to assess the condition of the Nation’s natural resources and contribute to decisions on environmental protection and management. Data collection and analysis were conducted by EPA Office of Research and Development (ORD) laboratory researchers, contractors, and cooperators following EMAP standards and protocols, and using EMAP sampling and analytical tools. EMAP–IM was conducted by Central EMAP–IM, a group and information system, guided by a strategic plan (Shepanek 1994) that provided for information management to support a national ecological monitoring program. Known, specific environmental parameters were collected with rigorous quality assurance and data documentation standards. The intent was for ORD to control sample collection, quality assurance, data processing and management, data analysis and assessment, and documentation through this information management system. The original EMAP–IM system included:

- a client-server Oracle database that included the EMAP Data Directory and data;
- EMAP Data Catalog (metadata files) in WordPerfect files;
- data sets in Oracle, SAS, and ASCII;
- an EMAP Gopher and web server; one of the first EPA programs to provide a public access web site; and
- automated data management and analysis functions for promoting data uses appropriate to the EMAP design (e.g., built-in tools for generating cumulative distribution functions and estimates of resource percentages exhibiting certain environmental conditions).

The early IM activities included a research component for determining the most effective system. Development of a robust information system for processing, managing, analyzing, and disseminating EMAP data was also a major objective of the program.

A bibliography of references relevant to the development of the EMAP Information Management System and Program is presented in Appendix L.

1.6 Current EMAP

The EMAP program was redirected by ORD in 1995 and now combines new monitoring with integration of data from research partners and historical sources. It involves partners from many agencies and institutions. Broad objectives of the program include:

- advancing the science of ecological risk assessment;
- promoting new approaches to monitoring in EPA Regions and Program Offices, and in state and local entities affiliated with these offices, through a program of smaller community-based projects;
- guiding national monitoring with improved scientific understanding of ecosystem integrity and dynamics;
- analyzing multi-scale data, aggregating among tiers, and integrating multi-resource data;
- supporting CENR goals for a national monitoring and research network;
- demonstrating the CENR framework in large regional-scale assessments; and
- helping to solve the scientific barriers to implementing the CENR framework.

The current EMAP program has the same mission as the earlier program but is reaching its goals in a very different way. It integrates both research and monitoring approaches, as well as data from

Section 1, Introduction and Approach

many sources (e.g., federal, state, local, academic) into ecological assessments. These assessments are led by EMAP Working Groups, whose membership includes (but is not limited to) cooperating research partners in EPA and non-EPA agencies and institutions. Non-EPA research partners include the National Oceanic and Atmospheric Administration (NOAA), U.S. Forest Service (USFS), U.S. Geological Survey (USGS), the U.S. Fish and Wildlife Service (USFWS), state government offices, academic researchers, and others. Research partners conduct studies to develop indicators and multi-tier sampling designs for monitoring the condition of ecological resources, and to field-test and apply the research in geographic studies. The planned studies cover a broader range of resource types and study focuses than the original program. These studies are outlined in the EMAP Research Plan (U. S. EPA 1997b) and in Working Group implementation plans. Section 2, EMAP Data, provides descriptions of data being collected in these studies.

EMAP's agenda is influenced by its participation in the federal interagency Committee on Environment and Natural Resources (CENR 1998a, 1998b, 1998c), which promotes the integration of environmental monitoring and assessment data from many sources to support assessment of regional and national trends in environmental quality. CENR is a committee of the White House Office of Science and Technology Policy's (OSTP) National Science and Technology Council (NSTC; a standing, cabinet-level body established by President Clinton in November 1993) whose membership includes a comprehensive group of Federal environmental and scientific agencies. CENR's mission is to develop effective responses to environmental problems through multi-disciplinary, interagency, and policy-relevant environmental research.

To support this mission, CENR is coordinating historically decentralized research programs and encouraging integration of their existing information systems into a network of distributed data maintained by different agencies. The focus of these efforts is to improve the availability and quality of inventories, surveys, and intensive monitoring, and to research and conduct natural resource management and environmental protection. One of CENR's goals is to adopt data standards and protocols for documenting data and facilitating exchange, such as: Z39.50, Federal Geographic Data Committee (FGDC) metadata standards, and the Global Change Research Program's (GCRP) Global Change Data and Information System (GCDIS). These activities allow CENR to oversee establishment of a national monitoring framework with common data standards for integrating and disseminating future environmental data.

EPA/EMAP's role in CENR is to participate on subcommittees developing the framework. This participation will shape EMAP's research and data management efforts because the EMAP-IM system will evolve in accordance with emerging CENR information management standards to ensure maximum interoperability of EMAP data with the data of other participating agencies.

1.7 EMAP Information Management

This section presents an overview of the EMAP–IM approach currently being implemented to support EMAP goals. Information management results both in a physical system infrastructure as well as software and databases designed to meet user needs for distributing information useful to both EMAP and other information users. In order to maintain the EMAP–IM system for these purposes, the EMAP information managers continue to take advantage of new computer and information management technology.

In the current program, information management is no longer controlled by a central group. EMAP–IM consists of a number of overlapping groups, including:

- **EMAP–IMWG**—including information management representatives from all Working Groups and ORD Laboratories involved in EMAP;
- **EMAP–IM (AED)**—consisting of staff at the EPA National Health and Environmental Effects Research Laboratory (NHEERL) Atlantic Ecology Division (AED), which has been given the lead for information management and has developed a new approach to serve the Working Groups. EMAP–IM (AED) programs the EMAP Public Web Site, maintains the EMAP Data Directory, and assists Working Groups with information management requirements;
- **EMAP–IM**—made up of all groups in EMAP that work on information management, including EMAP–IM (AED), information managers in the Working Groups and Resource Groups, and the IMWG.

Overall information management policies and direction are overseen by the IMWG and implemented by EMAP–IM (AED).

EMAP–IM’s plans are developed in accordance with EPA’s emphasis on the importance of information management, as demonstrated by the following references:

“EPA is committed to managing its information resources to provide the information necessary to inform and empower decision-makers to protect human health and the environment.” (U. S. EPA 1995b). Further, Goal 5 of the ORD Strategic Plan is “To provide reliable scientific, engineering, and risk assessment/risk management information to private and public stakeholders.” (U.S. EPA 1995a). Strebel and Frithsen (1995a) summarizes EMAP’s information management approach as follows:

“...EMAP inherits its charge to make data and information publicly available from legislative mandates to the US EPA to distribute data and information collected by the Agency. While providing data to users is an essential element of EMAP, this goal and the legislative mandates are reinforced by reviews of other information management systems developed and used by the Agency, the data policies adopted by the US Global Change Research Program, and the government emphasis on enhancing the electronic component of the Nation’s information infrastructure. National Research Council reviewers of EMAP have specifically encouraged the program to publish data using the Internet (NRC 1994).

...as the ability of [the Internet] to allow individual users access to vast information resources has been recognized, [its] use ... has grown dramatically. Information discovery tools such as Gopher and World Wide Web have converted an anarchy of individual anonymous File transfer Protocol (FTP) sites into an indexed and hyperlinked knowledge base. Academic use has shown the Internet to be an effective, if somewhat informal, publication medium. Government institutions are following this lead and commercial publishers are actively planning to add Internet offerings to their repertoire. It is abundantly evident that in the near future, the network will become a major repository and delivery system for information of all kinds. Preliminary demonstrations have shown how EMAP data and information could be accessed from the U.S. EPA Public Access Server by using commonly available information discovery tools such as Gopher, WAIS, and Mosaic.”

ORD is providing a broad framework of policies, strategies, and plans in its Strategic Plan (U. S. EPA 1997c) to guide the information support that EMAP needs. The ORD Science Information Management Coordination Board (SIMCorB) has also been formed to coordinate data and information management across ORD.

1.7.1 EMAP-IM Mission

Information management is a vital part of EMAP that supports program objectives by ensuring that EMAP data and resulting information are accessible and useful long beyond the initial studies that generated the data. The goal of EMAP-IM is:

“to support the EMAP mission by providing information management support to research efforts on monitoring and assessing the condition of the Nation’s ecological resources.” (U.S. EPA 1997a).

EMAP–IM objectives are intended to support the EMAP program objectives specified in the EMAP Research Plan (U.S. EPA 1997b) as shown in Table 1-2, below. Information management objectives at the level of the Working Groups is reviewed in Section 3, Information Management Needs and Requirements. Objectives include:

- Providing a Data Directory so that data of interest can be identified;
- Providing access to data and metadata files;
- Assisting with designing, developing, maintaining, operating, and/or deploying databases and access mechanisms for EMAP research activities;
- Providing IM support to EMAP Working Groups for planning, research, monitoring, and analysis efforts so that differences in information management environments of the groups are minimized;
- Ensuring a distributed data structure, allowing responsibility for the data to reside with the owners; and
- Maximizing interoperability with other environmental monitoring data systems in accordance with CENR objectives.

1.7.2 EMAP–IM Approach

To meet the objectives of the current EMAP program, the EMAP–IM system uses a more flexible approach that satisfies the growing need for EMAP researchers and managers to access and integrate data from EPA and non-EPA sources (Hale et al. 1998).

Current EMAP Working Groups collect and manage their own data with a greater variety of data collection, quality assurance, and data management methods than those used in the early EMAP program. EMAP does not take possession of all data sets, but relies on the data sources (Working Group researchers) for primary data processing and management, including quality assurance and metadata preparation. Each ORD Division or Working Group leading an EMAP project manages the data they collect or acquire and is responsible for quality assurance, documentation, and transfer to accessible Internet web sites. These groups are encouraged to follow established EMAP standards and formats (e.g., Frithsen and Strebel 1995; Strebel and Frithsen 1995a,1995b; NASA 1991, Appendix F of this Plan) in order to support data availability and to ensure the quality of data for integration across EMAP research organizations and with other environmental agencies and organizations (especially under CENR). The data collectors, in general, use their own data management and data analysis tools.

Section 1, Introduction and Approach

Table 1-2. How EMAP–IM Objectives Support EMAP Program Objectives

EMAP Program Objectives (EMAP Research Plan, U.S. EPA 1997b)	EMAP–IM Objectives
Advance the science of ecological risk assessment	<ul style="list-style-type: none"> • Improve mechanisms for data and metadata exchange among researchers
Promote new approaches to monitoring in EPA Regions and Program Offices and in participating state and local agencies through a program of smaller community-based projects in each region	<ul style="list-style-type: none"> • Develop a distributed database structure with external sources, allowing responsibility for the data to reside with the “owner” • Provide data directory, data, metadata, and web site
Guide national monitoring with improved scientific understanding of ecosystem integrity and dynamics	<ul style="list-style-type: none"> • Improve access to EMAP and non-EMAP data and metadata through Data Directory and EMAP Public Web Site, and coordinate efforts with other research organizations
Analyze multi-scale information, aggregate data among tiers, and integrate multi-resource data	<ul style="list-style-type: none"> • Improve access to EMAP and non-EMAP data and metadata through Data Directory and EMAP Public Web Site, and coordinate efforts with other research organizations • Provide information management support to EMAP Working Groups within ORD for planning, research, monitoring, and analysis efforts so that differences in the information management environment of the groups are minimized
Support federal interagency CENR goals for a national monitoring and research network	<ul style="list-style-type: none"> • Provide Data Directory compatible with Federal standards for interoperability • Provide Internet access to EMAP data sets for research and analysis • Maintain EMAP Public Web Site • Assist EMAP researchers in designing, developing, maintaining, operating, and/or deploying databases and access mechanisms for EMAP research activities
Demonstrate the CENR framework in large regional-scale assessments	<ul style="list-style-type: none"> • Ensure that required data and metadata are available for integration
Help resolve the scientific barriers to implementing the CENR framework	<ul style="list-style-type: none"> • Facilitate integration of data sets collected at different spatial and temporal scales • Maximize interoperability of EMAP tools (e.g., Data Directory) with other environmental monitoring data systems

To facilitate access to data that are physically stored at remote locations (e.g., universities, the EPA server at Research Triangle Park (RTP)), EMAP–IM system has evolved towards a model of federated databases with distributed data ownership and responsibility. Information management in the current program differs from that of the earlier program in that many current research partners are outside of EPA and may be obligated to follow standards very different from those maintained by the Program.

EMAP-IM uses a centralized index of widely distributed data consisting of the EMAP Data Directory and the EMAP Public Web Site. The cornerstones of this approach are available data, an index to locate data, a mechanism on the World Wide Web for accessing data, and high quality documentation for judging data usefulness. The successful delivery of data in this system depends on the use of uniform data standards and internal EMAP standards that ensure the quality and content of the Directory references, the data, and the metadata.

To successfully implement this distributed data model, EMAP-IM focuses its efforts in two main areas:

- **Maintain directories of data and documentation for EMAP users on the EMAP Public Web Site to ensure access to relevant data.** EMAP-IM's primary objective is to facilitate access to relevant data sets so they can be integrated into environmental assessments.
- **Participate in developing and adopting data standards that facilitate data integration into EMAP assessments and the CENR framework, and encourage their use by EMAP researchers.** All components of the IM system design evolved with emerging CENR standards. Where existing standards do not exist or are deemed to be inadequate, EMAP seeks venues for developing standards in accordance with CENR, International Standards Organization (ISO), and American National Standards Institute (ANSI) guidelines. In the absence of other options, EMAP-IM develops interim standards and facilitates review and maintenance of these standards. The standards adopted for use in the EMAP-IM system must be flexible, interoperable with other federal systems, and allow evolution with changing demands.

The EMAP-IM system outlined in this IM Plan is not a new implementation. The early EMAP-IM system components were designed to provide indexing, documentation, analysis, and data distribution. Modifications to the early EMAP system components are meeting the needs of the current EMAP program. A key subset of the original components have been upgraded to support expanded user needs, new data management techniques, and incorporation of new technology. For example, the Data Directory can be expanded to contain information about non-EPA research data sets used by the Working Groups; in the future, it will provide links to web sites where the data reside. The EMAP Public Web Site has been expanded to include links to data relevant to EMAP research (e.g., Multi-Resolution Land Characteristics data). The major components in the existing system that have been upgraded are briefly described in Table 1-3. For more detailed information, see Section 4, Technical Design.

Section 1, Introduction and Approach

Table 1-3. Existing EMAP Information Management System Components

Component	Description
Data Directory	The EMAP Data Directory is an index of distributed data, documentation, and information products relevant to EMAP users. The Data Directory format is based on established EPA guidelines (Strebel and Frithsen 1995b; updated in U.S. EPA 1996c). The Data Directory now exists as an Oracle database on the EMAP internal and public Web Sites. In the future, the Data Directory may provide hyperlinks to distributed sites where data and metadata are housed (e.g., World Wide Web pages, FTP sites, data repositories (e.g., STORET, ² or other sites). The Data Directory contains links to corresponding data and metadata files.
Data Catalog	The EMAP Data Catalog provides a format for housing EMAP metadata. Currently, it consists of text files containing detailed documentation of data sets from EMAP Resource Groups, R-EMAP, and related groups. It describes the entities (fields or columns in the database), the quality, the methods used, and other detailed information. Data Catalog files are made available from the EMAP Public Web Site.
EMAP Public Web Site	The EMAP Public Web Site provides web browser access to EMAP data and information for all users. The Web Site houses the EMAP Data Directory, Data Catalog files, data sets that have passed EMAP Quality Assurance requirements and are ready for public distribution, and program descriptions. The Web Site is housed on the EPA Public Access Web Server at RTP. Materials to be placed on the site must go through an approval process (Strebel and Frithsen 1995a; U.S. EPA 1998b).
EMAP Internal Web Site	The EMAP Internal Web Site is intended for testing of applications and review of data and information that are being prepared for the EMAP Public Web Site. It is housed on the EPA internal wide area network (WAN) at AED. It allows EPA researchers to access data in a secured network. The site contains the Oracle Data Directory database; EMAP documents and bibliographic references; Wide Area Information Servers (WAIS) searching; NASA Directory Interchange Format (DIF) authoring tool for preparing Data Directory entries; data sets being prepared for public release; and data, metadata, and Data Directory status information.

² EPA Storage and Retrieval of U.S. Waterways Parametric Data.

This IM approach fulfills the following needs:

- organizing and summarizing programs and data sets so they are easy to locate and understand;
- allowing data collectors to maintain the integrity and official versions of data sets;
- making metadata accessible to users;
- notifying users of data set corrections and updates;
- linking to non-EMAP web sites in order to obtain useful data;
- facilitating inquiries about Resource Group data sets; and
- listing and distribution of publications and program reports.

This system design forms the foundation of information management systems developed to support the ORD Regional Assessment pilot programs (e.g., MAIA, Western Pilot).

The model outlined in this Plan of raw field data and experimental data being managed by the data originators at distributed sites and summary data being distributed on the Internet is widely used in federal and state agencies, and in the scientific data community. Existing and emerging technologies and standards can be adopted to enhance this approach as needs evolve within EMAP. The discussions at Requirements Analysis meetings conducted for this version of the IM Plan indicated that researchers and partners anticipate a need for including relational or object-oriented database capabilities within this system to query complex monitoring data and aggregated results. The need for these capabilities is addressed in Appendix D, Preliminary Design and Options Document.

1.7.3 Role of EMAP-IM (AED)

EMAP-IM (AED) provides the following services to support the EMAP management and Working Groups:

- maintain EMAP Data Directory, Data Catalog, and data files;
- program and maintain web pages for the EMAP Public Web Site and Internal Web Site including the Data Directory, Catalog, data sets, and documents;
- coordinate information management efforts among individual Working Groups and the common EMAP information resources (including the Data Directory, EMAP web sites, data, metadata, and publications);
- develop and maintain EMAP data standards, procedures, and formats; and
- assist Working Groups with preparing and distributing EMAP data and documentation.

EMAP-IM (AED)'s role is the network coordinator for all groups. Working Groups generate and maintain data and forward information to EMAP-IM (AED) to be linked to or placed on the EMAP Public Web Site. For more information on EMAP-IM's project management structure, see Section 5, Project Management and Coordination.

1.8 EMAP Data Policy Statement

Acknowledgment: This Data Policy Statement was modified, with permission, from two sources:

Data Management for Global Change Research. Policy Statements for the National Assessment Program. July 1998. U.S. Global Change Research Program. National Science Foundation, Washington, DC.

Section 1, Introduction and Approach

U. S. GLOBEC. 1994. U. S. GLOBEC Data Policy. U. S. Global Ocean Ecosystems Dynamics. Report No. 10. Woods Hole, MA. (<http://globec.whoi.edu>).

The fundamental objectives of U.S. EPA Environmental Monitoring and Assessment Program (EMAP) are dependent upon the cooperation of scientists from several disciplines. Our objectives require quantitative analysis of interdisciplinary data sets, and therefore data must be exchanged between researchers. To extract the full scientific value, data must be made available to the scientific community on a timely basis.

Precedent and perception have resulted in a disparity of data collection, storage, and archival methods. This makes the exchange of data difficult and may suppress dissemination of data. EMAP seeks to enhance the value of data collected within the Program by providing a set of guidelines for the collection, storage, and archival of these data sets.

The overall purpose of these policy statements is to facilitate full and open access and use with confidence, both now and in the future, of the data and information that is used in and results from the Environmental Monitoring and Assessment Program activities. These policies reflect the goals and policies of EMAP and incorporate federal laws, directives, and regulations regarding the maintenance and dissemination of data and information in the Federal Government. They apply to all participants in the Environmental Monitoring and Assessment Program, including federal, state, local, tribal, foreign, educational, and non-government organizations and their private partners.

- The Environmental Monitoring and Assessment Program requires a continuing commitment to the establishment, maintenance, description, accessibility, and long-term availability of high-quality data and information.
- Full and open sharing of the full suite of data and published information produced by the Environmental Monitoring and Assessment Program is a fundamental objective. Data and information should be available without restriction for no more than the cost of reproduction and distribution. Where possible, the access to the data should be via the World Wide Web to keep the cost of delivery to a minimum and to allow distribution to be as wide as possible.
- Organizations and individuals participating in the Environmental Monitoring and Assessment Program should make measurements which do not involve manual analysis available to other program participants within 6 months after collection. All other measurements should be made available to program participants within 12 months after collection. Data and metadata should be publicly available on the Internet within 18 months after field collection. These are target goals; advise the Chair of the Information Management Working Group if they cannot be met.

Section 1, Introduction and Approach

- All data sets and published information used in the Environmental Monitoring and Assessment Program should be identified with a citation; for data sets an indication of how the data may be accessed should be provided.
- National and international standards should be used to the greatest extent possible.
- All data sets generated as part of the Environmental Monitoring and Assessment Program must be described and a quality assessment provided. All such data set descriptions should be made available for inclusion in the EMAP Data Directory and Data Catalog. In addition, steps should be taken to assure their continuing availability. Spatial data set descriptions should be compatible with the Content Standards for Digital Geospatial Metadata of the Federal Geographic Data Committee.
- Organizations and individuals participating in the Environmental Monitoring and Assessment Program should actively participate in its Web page to share information and coordinate the Program's disparate activities. The identifications of all the Program's published information, data sets, and data set descriptions should be made accessible over the Internet.
- Publication of descriptive or interpretive results is the privilege and responsibility of the investigators who collect the data, as is the publication of high-quality data sets for use by others. 'Investigator' means any participant (ORD, Region, state, tribe) who plays a role in data collection. The purpose of data exchange is to facilitate collaboration between scientists, the combination of multiple data sets for interdisciplinary and comparative studies, and the development and testing of new theories. Any person making substantial use of a data set within 18 months after field collection should communicate with the investigators who acquired the data and give proper attribution or co-authorship. After a dataset is moved into the public domain, there are no restrictions except to use the suggested data set citation.
- Requests for exemptions from these data policy statements should be submitted to the EMAP Director.
- Suggested Data Product Requirement for Grants, Cooperative Agreements, and Contracts: *Describe the plan to make available the data products produced, whether from observations or analyses, that contribute significantly to the <grant's> results. The data products will be made available to the <grant official/contracting officer> without restriction and be accompanied by comprehensive metadata documentation adequate for specialists and non-specialists alike to be able to not only understand both how and where the data products were obtained but adequate for them to be used with confidence for generations. The data products and their metadata will be provided in a <standard> exchange format no later than the <grant's> final report or the publication of the data product's associated results, whichever comes first.*

Section 1, Introduction and Approach

AUTHORITIES AND REFERENCES. As reflected in the following authorities and references, the Executive and Legislative branches of the U.S. Government both recognize the need for federal agencies to assume an active role in providing information to the public.

- a. *Privacy Act of 1974* restricts the government's ability to disseminate information that could invade the personal privacy of an individual. Privacy Act data cannot be released without appropriate review.
- b. *The Freedom of Information Reform Act (FOIA) of 1986* establishes what agencies must make available to the public in terms of public information, agency rules, opinions, orders, records, and proceedings.
- c. *OMB Circular No. A-16, Coordination of Surveying and Mapping Activities*, October 19, 1990, establishes coordination procedures for federal agencies and describes the responsibilities with respect to coordination of those federal surveying and mapping activities.
- d. *Executive Office of the President, Data Management for Global Change Research Policy Statements* issued in July 1991 provided a set of policy statements to facilitate full and open access to quality data for global change research.
- e. *Land Remote Sensing Policy Act of 1992*, requires that unenhanced data from Landsat 7 and other government-funded and -owned land remote sensing systems be made available to users at the cost of fulfilling user requests and on a non-discriminatory basis.
- f. *The White House Memorandum on the Administration of the Freedom of Information Act (FOIA)* issued October 4, 1993, states that a commitment to openness requires more than merely responding to requests from the public. Each agency has a responsibility to distribute information on its own initiative, and to enhance public access through the use of electronic information systems.
- g. *Executive Order 12862, Setting Customer Service Standards*, September 11, 1993, mandates easy accessibility of federal government information and services.
- h. *OMB Circular No. A-130, Management of Federal Information Resources*, June 25, 1993, states that every agency has a responsibility to inform the public within the context of its mission. This responsibility requires that agencies distribute information at the agency's initiative, rather than merely responding when the public requests information.

Section 1, Introduction and Approach

- i. *Government Performance Results Act (GPRA) of 1993* requirements are intended to improve federal program effectiveness and public accountability by promoting a focus on results, service quality and customer satisfaction.
- j. *44 United States Code Chapter 31 - Records Management by Federal Agencies* requires agencies to create and maintain documents and provides the basis for public records and information.
- k. *44 United States Code Chapters 17 and 19* define the legal requirements for providing information to the public through the Federal Depository Library Program.
- l. *Executive Order 12906, Coordinating Geographic Data Acquisition and Access; The National Spatial Data Infrastructure*, April 11, 1994, requires each agency to document all new geospatial data it collects or produces, either directly or indirectly, using the developing FGDC standard, and to make that documentation electronically accessible.
- m. *U.S. Environment and Natural Resource Data Access System Guideline*, July 6, 1995, requires all federal agencies participating in environment and natural resources research to develop their data and information search and access systems to have at least Internet connectivity and be ANSI Z39.50 compliant.
- n. *Paperwork Reduction Act (PRA) of 1980, as amended 1995*, requires agencies to provide for the dissemination of public information on a timely basis, on equitable terms, and in a manner that promotes the utility of the information to the public and makes effective use of information technology.
- o. *Electronic Freedom of Information Act (EFOIA) of 1996* mandates that agencies make all reasonable efforts to provide information available to requesters in the medium of their choice.
- p. *OMB Bulletin 98-5, Establishment of Government Information Locator Service (GILS)*, February 6, 1998, is designed to help the public and agencies locate and access information electronically throughout the U.S. government.

1.9 Conclusions

The Internet-based system outlined above facilitates access to geographically dispersed data sets for EMAP researchers and a wide range of potential users. This multiple-agency, cooperative approach to data dissemination and exchange has prompted enhancements to the original EMAP system that increase the flexibility of the system and distribute the responsibility for data collection as well as data management and ownership. This information management approach serves current program needs and is flexible enough to respond to future program needs. For example, the system could be expanded to respond to the following potential EMAP needs:

- relational or object-oriented databases could be developed if there were a need for increased centralization of data management and storage;
- web server technology could be incorporated to increase direct access to distributed data sets; or
- new configurations and functionality could be implemented to expand the interoperability of components such as the Data Directory with those of other agencies.

Some of these needs are already being addressed or may be further expanded in the future (see Section 6, Implementation Plan and Appendix D, Preliminary Design and Options Document).

Section 2 EMAP Data

- 2.1 Introduction**
- 2.2 EMAP 1990–1995 Data**
- 2.3 Current EMAP 1996–2001 Data**
 - 2.3.1 ORD Regional–Scale Assessments Program—Mid-Atlantic Integrated Assessment and Western Pilot**
 - 2.3.1.1 Mid-Atlantic Integrated Assessment—Estuaries**
 - 2.3.1.2 Mid-Atlantic Integrated Assessment—Surface Waters**
 - 2.3.1.3 Mid-Atlantic Integrated Assessment—Landscape Ecology**
 - 2.3.2 Intensive/Index Sites**
 - 2.3.2.1 Demonstration Intensive Sites Project**
 - 2.3.2.2 Coastal Intensive Sites Network (CISNet)**
 - 2.3.3 Landscape Ecology**
 - 2.3.4 Regional EMAP**
 - 2.3.5 Ecological Indicator Development**
 - 2.3.5.1 Aquatic Mortality Monitoring Database**
- 2.4 Conclusions**

EMAP Working Groups are participating in a wide variety of data collection and analysis projects that are being conducted by EPA and non-EPA researchers. These partners share in the collection, management, and delivery of data to potential users, but the varying project management structures of each Working Group means that no two groups will complete the task in the same way. Because Working Groups perform most data management and delivery tasks themselves, the EMAP-IM system is made up of data sets that are distributed at many sites. The scope of the data being created through these partnerships is described in this section, along with the management practices and planned distribution mechanisms for the data.

2.1 Introduction

This section provides an overview of EMAP data collection and management activities. Additional detail about the types, volumes, status, and repositories for the data and documentation is provided in Appendices B (Data Management Needs and Practices of EMAP Working Groups) and C (Inventory of EMAP Data), and the EMAP Public Web Site (EMAP 1998).

There are different kinds of data. Aggregate data are statistical summaries and data derived from modification of the original data through analysis, integration, and enhancement to produce a new data set; they can be created from data collected in different regions or at different times. Examples of data aggregates include:

- integration of parameters from sample replicates at one station;
- a “benthic index” that integrates measurements of dissolved oxygen, salinity, inorganic concentrations, and benthic abundance;
- results of analysis of raw data sets.

Summary data represent the results of analysis or data collection that places the data into a context (e.g., aggregate or raw data from all stations sampled summarized as percentages of the total number).

Preliminary data are those under development and not yet ready for public release. Completed data are those that have gone through extensive quality assurance and review procedures and been approved for public release.

2.2 EMAP 1990–1995 Data

Early EMAP program data and information are an important source of historical information and analytical tools for Working Groups. The majority of 1990–1995 data sets are managed within ORD laboratory data management systems. Data sets include raw field sampling data, aggregate data, and summary data created as a result of analyses and data integration.

Each Resource Group handled data largely by its own standards, although there was some cross-group standardization on codes, metadata, and other issues. Some Resource Groups centralized data management at a single research site, and others allowed research collaborators to manage data at independent sites. For example, the EMAP–Estuaries Resource Group database consists of raw data collected and analyzed by regional groups who transferred the data aggregates and summaries to a central database. Raw data were handled at three separate locations and by two agencies (EPA and NOAA). In Resource Groups such as EMAP–Surface Waters, data were collected and managed

Section 2, EMAP Data

by a single organization that stored both raw data and data aggregates at the same location. EMAP–Forests and EMAP–Agroecosystems Resource Group data are being handled by agencies other than EPA (see Table 2-1).

The Central EMAP–IM architecture from the early EMAP program was designed as an Oracle system with a central node and distributed nodes at each Resource Group site. This system was never fully implemented. The EMAP–Estuaries Resource Group had three years of data for the Virginian Province loaded into the central system, some EMAP–Surface Waters data were loaded, and links were provided to the EMAP–Forests database at Las Vegas. Data sets from other Resource Groups were not loaded (e.g., Agroecosystems, Forests, Rangelands).

Raw data from the early EMAP program remains the responsibility of the Resource Groups (now the NHEERL and NERL laboratories) that collected them. Each group that collected the data is maintaining the data on its own computers using existing standards for that organization. All systems created and used for field sampling, preliminary data processing, quality assurance/quality control (QA/QC) and data analysis are maintained at the discretion of the originating group.

The repositories (data storage and distribution locations) of the raw and summary data for each Resource Group are shown in Table 2-1. These summary data are available on the EMAP Public Web Site. Raw data are maintained by the researchers and can be obtained upon request. For information about the types and status of individual data sets, see Appendix C, Inventory of EMAP Data, and the EMAP Public Web Site (EMAP 1998).

Section 2, EMAP Data

Table 2-1. Repositories of 1990–1995 Raw and Summary Data Sets

Resource Group	Location of Raw Data	Location of Summary Data
Estuaries		
Virginian Province	ORD-AED	EMAP Public Web Site
Carolinian Province	NOAA at the South Carolina Marine Resources Research Institute	EMAP Public Web Site
Louisianian & West Indian	ORD-Gulf Ecology Division (GED)	EMAP Public Web Site
Surface Waters	ORD-Western Ecology Division (WED)	EMAP Public Web Site
Wetlands	ORD-WED	ORD-WED
Forests	Data management at U.S. Forest Services (USFS) Forest Health Monitoring (FHM) Program Data stored at ORD-Landscape Ecology (Environmental Sciences Division, NERL, Las Vegas, NV)	Data management at USFS FHM Program Data stored at EPA Landscape Ecology Branch, Environmental Sciences Division, NERL
Rangelands	ORD-Landscape Ecology	ORD-Landscape Ecology
Agroecosystems	U.S. Dept. of Agriculture (USDA; access restricted by law)	North Carolina State University, USFS, and EPA-NERL
Great Lakes	ORD-Mid-Continent Ecology Division (MED)	EMAP Public Web Site
Landscape Characterization	ORD-Landscape Ecology (Environmental Sciences Division, NERL, Las Vegas, NV)	ORD-Landscape Ecology (Environmental Sciences Division, NERL, Las Vegas, NV) and EMAP Public Web Site

For more information, see the EMAP Public Web Site program contacts and data status areas.

2.3 Current EMAP 1996–2001 Data

In the current program, each Working Group includes a number of EPA and non-EPA research partners (from many different agencies, research laboratories, and other organizations) who are responsible for managing, documenting, and distributing EMAP data. These partners are listed in Table 2-2.

Section 2, EMAP Data

Table 2-2. Working Group Partners Responsible for Data Sets

Working Group	Data Partners
ORD Regional Assessments/ Mid-Atlantic Integrated Assessment (MAIA)–Estuaries group ORD-AED, lead	ORD-AED Chesapeake Bay Program (CBP) NOAA–National Status and Trends Program - Delaware Bay NOAA–Charleston Delaware River Basin Commission National Park Service (NPS)–Maryland Coastal Bays Monitoring EPA Region III
ORD Regional Assessments/ MAIA–Surface Waters ORD–WED, lead	ORD-WED Various EPA and academic scientists who serve as Indicator Leads EPA Region III
ORD Regional Assessments/ MAIA-Landscape Ecology ORD-NERL Laboratory Landscape Ecology Branch, lead	ORD-Landscape Ecology (Environmental Sciences Division, NERL, Las Vegas, NV) Tennessee Valley Authority (TVA) U.S. Geological Survey (USGS) EPA Region III MRLC
Intensive Sites /Demonstration Intensive Sites Project (DISPro) ORD-WED, lead	ORD-WED NPS UVB EPA NERL
Intensive Sites /Coastal Intensive Sites Network (CISNet) ORD-GED and NOAA–National Status and Trends Program, leads	Grant recipients
Landscape Ecology ORD-NERL Landscape Ecology Branch, lead	ORD–Landscape Ecology (Environmental Sciences Division, NERL, Las Vegas, NV) TVA USGS
R-EMAP ORD-MED, lead	All ten EPA Regional Offices and their subcontractors (states, academic scientists, regional research boards)
Ecological Indicators ORD-GED, lead	ORD, academic, and other researchers developing indicators independently (Aquatic Mortality database has State agency partners)

Types of data and information products include:

- monitoring data sets;
- aggregate data sets resulting from integration of existing data into assessments (e.g., value-added data sets, which include results of applying developed indices; community-level summaries using established methodologies);
- documentation of data sets (metadata) in EMAP Data Directory, Data Catalog, FGDC, or other formats. Metadata includes how data were collected, by whom, under what

Section 2, EMAP Data

- sampling design, with what types of instruments, methods used, QA/QC applied, statistical algorithms used, analytical tools and indices, and assumptions made in data analysis, etc. (e.g., statistical methods applied, indices and their derivations);
- standards (e.g., documentation);
 - tools for evaluating environmental condition (e.g., diversity indices, composite indicators);
 - methodology for evaluating data (e.g., methodologies for integrating random and non-random sampling designs, spatial statistics applicable to landscape ecology);
 - reports (e.g., statistical summaries, regional resource assessments);
 - spatial data sets (e.g., remote sensing images, geographic information system [GIS] data sets, landscape indicators); and
 - maps, charts, graphs, data tables, and other derived products.

EMAP-IM (AED) does not take physical possession of most of the data and information products from the Working Groups, except for those directly managed by AED (e.g., MAIA-Estuaries data). Instead, data sources (i.e., research partners who collect and create data sets) use their own data management systems and standards to maintain and distribute raw and aggregate data at their sites in many different formats. EMAP-IM (AED) tracks relevant data at the independent sites. To ensure the success of the tracking system, the data sources (researchers who collect, maintain, and distribute the data) take responsibility for appropriate data management procedures (e.g., QA/QC and preparation of metadata), documentation of data characteristics and history, and adequate distribution of data sets. EMAP-IM (AED) can provide assistance to EMAP researchers to ensure successful completion of these tasks.

The following subsections summarize the status of EMAP Working Group data that are tracked. Working Groups are in different stages of work. Some are already generating data, while others are still formulating research plans. The information presented here was collected by reviewing the EMAP Research Plan (U.S. EPA 1997b) and conducting Requirements Analysis interviews with the Working Groups. More detailed information about data activities of the Working Groups can be found in Appendix B, Data Management Needs and Practices of EMAP Working Groups, and the EMAP Public Web Site. Estimates of data volumes and status are contained in Appendix C, Inventory of EMAP Data. Access to completed summary data is through the EMAP Public Web Site. This Information Management Plan will be updated in the future to include new information about these activities.

Section 2, EMAP Data

Table 2-3. Location of Raw and Summary Data for Working Groups

Working Group	Location of Raw Data	Location of Summary Data
ORD Regional Assessments– MAIA–Estuaries	ORD–AED	EMAP Public Web Site
MAIA–Surface Waters	ORD–WED	EMAP Public Web Site
MAIA–Landscape Ecology	ORD–Landscape Ecology Branch	EMAP Public Web Site MAIA web site
Western Pilot	Same as MAIA	EMAP Public Web Site, STORET
Intensive Sites–DISPro	EPA EIMS NPS Air Resources Division University of Georgia National Ultra–Violet Monitoring Center NUVMC)	EPA EIMS NPS–Air Resources Division Canadian World Ozone and UV Radiation Data Centre (WOUDC)
Intensive Sites–CISNet	ORD–GED	NOAA ORD–GED
Landscape Ecology	ORD–Landscape Ecology Branch	Surf Your Watershed EMAP Public Web Site
R–EMAP	Universities, states, regional planning boards, EPA Regional Offices	Universities States Regional planning boards EPA Regional Offices STORET ¹ (water, fish, sediment data) Mercury Deposition Network (MDN; air data) Surf Your Watershed (landscape ecology data) EMAP Public Web Site Others to be determined
Ecological Indicators	ORD–GED	EMAP Information Management System, individual researchers' databases

¹ STORET is the EPA Storage and Retrieval of U.S. Waterways Parametric Data.

Repositories for these data sets are further described in the sections for each Working Group, below. Any web-accessible site linked to the EMAP Public Web Site can serve this purpose. One repository that will be used by Working Groups, especially R-EMAP, is the modernized STORET (STORET 1998). Modernized STORET is popular among the R-EMAP groups because the states, who are their research partners, are required to enter these monitoring data into it so it can be aggregated with data sets from other programs. STORET will provide capabilities for many types of analyses, for documentation of the data, data entry tools, and report production tools. These capabilities are useful to EMAP groups and will allow them to use and store their data in a standardized system accessed

by many users. STORET will have fields for identifying the data as R-EMAP or EMAP so it can be segregated from other data.

2.3.1 ORD Regional-Scale Assessments Program—Mid-Atlantic Integrated Assessment and Western Pilot

The EPA ORD Regional-Scale Integrated Assessment Program is an effort by EMAP and CENR (CENR 1998a, 1998b) to resolve fundamental scientific issues involved in the integration of existing data from different sources with new data for regional assessments and decision-making regarding environmental resources. Although EMAP is designed to be implemented on several spatial scales, its primary goal is to assess the status and trends of ecological resources on regional and national scales. The Regional-Scale Integrated Assessments Program offers EMAP the opportunity to implement its design in a single ecological region and to address integration issues at different spatial scales. Data and information are used in the integrated assessments to develop and field-test methods for conducting regional ecological assessments, implement a regional-scale assessment design for a single ecological region, and address issues of data integration at different spatial scales.

To accomplish this goal, the Regional Assessments are designed to identify, organize, and analyze pertinent data, information, and tools, and facilitate the access and integration of these products into a framework that supports use of the information in decision-making processes designed to manage regional resources. The purpose of these activities is to conduct “integrated assessments” in which data from relevant sources are combined to assess the status of selected resources (e.g., streams) and regions (e.g., watersheds).

EMAP has chosen MAIA for the first regional assessment. The region includes the land area and near-coastal waters for all of PA, WV, MD, DE, VA, and parts of NJ, NY, and NC. It has been chosen as the pilot because extensive data sets have been collected in this region by numerous EMAP and non-EMAP research and monitoring programs. Several EMAP Resource Groups have also conducted demonstration projects in the Mid-Atlantic region. The synergy of these programs and data provides an opportunity to draw upon diverse outside expertise and a wealth of monitoring data from a variety of federal (multi-agency), EPA Regions, state, academic, local, and private programs. MAIA also serves as a pilot for the CENR Environmental Monitoring Steering Committee to evaluate the CENR monitoring framework (U.S. EPA 1996a). EMAP and CENR will use the lessons learned to conduct assessments in other U.S. regions after MAIA is completed.

MAIA integrated assessments are focused in three areas: estuaries, streams, and landscape ecology. These three areas are reviewed in more detail in the following pages. The MAIA assessment process will result in a set of data and information products that provide a legacy for use by local, regional, and national researchers and resource managers to assist in the design and implementation of future monitoring programs to fill critical data gaps. MAIA will deliver the following information products:

Section 2, EMAP Data

- inventory of environmental monitoring and research data collection programs which documents the types and availability of data sets from these programs and analyzes their usefulness relative to the MAIA program;
- directory of available data sets;
- EMAP 1990–1995 MAIA region data sets;
- state, federal, and other regional data specified in the MAIA implementation plan;
- EPA State of the Environment reports (status and trends of resources);
- landscape ecology atlas of environmental condition in the Mid-Atlantic region;
- integrated assessment analysis results (data sets including derived indicators and index results, interpretation of raw data, integrated results, indicators and indices, and conclusions);
- methodology (assessment techniques, algorithms, tools, indicators) for integrating data from different spatial scales and sampling designs for conducting integrated assessments of resources, designing monitoring programs, and managing data; and
- identification of data gaps and evaluation of their importance in managing the regional resources.

MAIA is being conducted as a partnership between ORD, EMAP, and EPA Regions II, III, and IV to support development of the technology needed to address assessment questions of importance to environmental and resource managers. The partnership allows Region III to provide EMAP with client-based feedback about the utility of assessment results, and for EMAP to gain access to additional sources of regional data through Region III's continuing regional assessment efforts.

Starting in FY1999, EMAP will begin implementing a Western pilot assessment project. This effort will be similar to MAIA and will test the transfer of the EMAP approaches to another region. Approaches developed by EMAP and previously implemented and enhanced by MAIA will be instituted in cooperation with EPA Regions VIII, IX, and X. Data management approaches will be tested and refined in this new regional program.

2.3.1.1 Mid-Atlantic Integrated Assessment–Estuaries

The purpose of MAIA–Estuaries data collection is to estimate the ecological condition of mid-Atlantic estuaries and produce a “State of the Estuaries” integrated assessment for the Mid-Atlantic region. Data collection and analysis are being coordinated by the EMAP-IM (AED) and conducted by a consortium of coastal monitoring organizations (see Table 2-4) that have been involved in long-term field monitoring in this region.

Section 2, EMAP Data

Table 2-4. MAIA–Estuaries Data Management

	MAIA–ESTUARIES
Data Collection & Existing Sources	Data collection began in 1997 and is continuing in 1998. It includes the same parameters as those collected by the EMAP–Estuaries Resource Group with the addition of nutrients and some toxics (including water quality parameters, benthic infauna, sediment toxicity and chemistry (1997 only), and fish trawls (1998 only)). Sampling teams use bar-coded sample ID’s. However, these tools are only being used by AED, GED, and NPS; other groups used their own systems which are mostly manual.
Data Aggregates & Products	Integrated assessments are being produced by the MAIA Analytical Team (including all research partners).
Georeferencing and GIS Products	AED is creating GIS data in the form of station locations that can be combined with commonly available GIS coverages of the MAIA region. They are also using USGS Digital Raster Graphs (DRG’s) to develop watershed boundaries for the estuaries they sampled. In addition, they use Landscape Ecology indicator coverages developed for the MAIA landscape atlas with their sampling data for the integrated assessments.
Data Integration Issues	The MAIA Analytical Team is incorporating summary data from all collaborators into an analytical database for the integrated assessment. In order to do this, AED researchers must reconcile differences among data collected by different researchers. For example, some stations were located outside the EMAP sampling grid, so the MAIA–Estuaries group needs to match research collaborators’ station names with EMAP station names, add the extra stations, and generate inclusion probabilities. This problem is an ORD research question about how to integrate probability-based analysis, fixed stations, and remote sensing. ORD will publish an analysis of this research problem.
Methods, Algorithms, Models, Equations, & Indicators	<p>AED will continue to refine the Benthic Index developed in the early EMAP program, and may also develop a Fish Index. They will use landscape indicators developed by Landscape Ecology (see Section 2.3.3, Landscape Ecology, below) to determine associations with estuarine indicators.</p> <p>AED is also producing “integrated assessments” of ecological indicators by comparing ecological indicators to one another. For example, they are now comparing the EMAP Benthic Index with the Chesapeake Bay Benthic Index on a site-by-site basis. Other indicators may be compared in the future as appropriate questions are posed for research.</p>
Data Management, QA/QC, Standards, & Long-Term Maintenance	The collaborating groups that collect data are responsible for QA/QC, management, and analysis of the data. These groups submit aggregate (summary) data sets to EMAP-IM (AED). In order to conduct the integrated assessments, AED data analysts re-format and load the data into the internal MAIA Estuaries SAS database for use by EPA/AED scientists.

Section 2, EMAP Data

Table 2-4. Continued

	MAIA-ESTUARIES
Data Management, QA/QC, Standards, & Long-Term Maintenance continued	<p>EMAP-IM has established data submission, content, and format standards for contract analytical laboratories (Buffum and Hale 1997a, Buffum and Hale, 1997b). These standards were reviewed by the collaborating groups, who may or may not use them. These groups continue to do database design and management according to their own needs and capabilities.</p> <p>The EMAP-IM (AED) will provide long-term stewardship for summary data collected and generated by AED in field sampling, laboratory analysis, data aggregation, and integrated assessments. They will also provide documentation for these products. Version control is now being done on data sets submitted to AED by keeping one Read-Only directory on the internal AED network server for final versions of data sets where EPA internal web site users can access them.</p>
Data Distribution	<p>The summary database will be made publicly accessible on EMAP/MAIA public web sites; original data will reside with the collecting organizations, such as the CBP (CBP 1998). The EMAP Public Web Site will link users to these sites. Data will be in a variety of formats, including SAS and Arc/Info export files.</p> <p>The group needs to exchange preliminary data among researchers, but cannot use the EPA internal web servers for this task because non-EPA researchers (e.g., NOAA) cannot access this network. Preliminary data cannot be posted on the public web site because they have not been approved for public release. In the short-term, the goal is to distribute preliminary data to research partners via email attachments, FTP, or diskette, and in formats responsive to group needs, such as SAS, ASCII format (comma-separated values/CSV). These obstacles to data sharing need to be addressed because collaborators need to review documents and data before they are published. EPA could provide a solution through a proposed "extranet" (see Section 5.6.8, Data Exchange among EMAP Researchers) or through options that are proposed in the future by SIMCorB².</p> <p>At the end of the MAIA program, AED will distribute the finished database (summary data and analytical tools) to groups in the MAIA region for their use in managing resources. EMAP-IM (AED) and the other data collection groups (e.g., Chesapeake Bay) are the long-term stewards for their data, but regional groups maintain some responsibility for maintaining and distributing summary data, results of the integrated assessments, methodology, and standards.</p>
Data Documentation	<p>AED will create entries for the EMAP Data Directory and Data Catalog for data sets it collected. A subset of EMAP Data Catalog standards was distributed to all participants for use in preparing documentation. Collaborators have been asked to follow these standards, but most researchers are still analyzing data so the documentation is not yet completed.</p>

² SIMCorB is a new group that coordinates data management activities within ORD (see Section 5.5.1, for more information).

Section 2, EMAP Data

2.3.1.2 Mid-Atlantic Integrated Assessment–Surface Waters

The purpose of data collection and analysis in this group is to estimate the ecological condition of mid-Atlantic streams by conducting status and trends assessments for MAIA streams and to create watershed metrics or indices that aggregate stream characteristics (physical, chemical, biological) in order to characterize watershed condition.

Table 2-5. MAIA–Surface Waters Data Management

	MAIA–SURFACE WATERS
Data Collection & Existing Sources	<p>ORD’s WED coordinates collection of stream monitoring data by EPA and non-EPA field and lab teams. Raw data include thousands of parameters collected from 150–200 stream stations every summer from 1993–1997; data collection is continuing in 1998. These data are transferred to ORD, who forward the data to researchers known as “Indicator Leads.” These researchers coordinate development of ecological indicators (or metrics) that are used to measure ecological condition in the surface waters (see Methods, Algorithms, Models, Equations & indicators box, below).</p> <p>MAIA–Surface Waters also relies on a number of non-EMAP data sources for their analyses. One data set critical to their efforts is the EPA River Reach (RF3) database. The quality of these data varies widely by region, although the data in the Region III portion of the MAIA study area had recently been updated. Surface Waters also uses Natural Resources Conservation Service’s (NRCS) Natural Resources Inventory (NRI) (NRCS, 1998) data, and the Multi-Resolution Land Characteristics (MRLC) land cover data.</p>
Data Aggregates & Products	Data aggregates will be produced in the process of creating watershed metrics by combining the field and laboratory data into data aggregates with other data sources (e.g., land cover from Landscape Ecology, road networks from U.S. Census TIGER data, point sources from the EPA National Pollutant Discharge Elimination System (NPDES) database, or locations of mine drainage from previous independent studies).
Georeferencing and GIS Products	GIS coverages will be produced for assessments.
Data Integration Issues	There will be a need to reconcile species codes between this database and the USGS National Water Quality Assessment (NAWQA) database, since USGS is using Integrated Taxonomic Information System (ITIS, 1998) species codes and the Surface Waters group is not currently using them.
Methods, Algorithms, Models, Equations, & Indicators	<p>The metrics created by this Working Group are an aggregation of the raw field and lab data about watershed stressors or landscape condition (such as number of miles of roads in a watershed or the distribution of substrate types along a particular stream reach). These metrics are created by combining the field/lab data into data aggregates with other data sources (such as land cover from Landscape Ecology, road networks from U.S. Census TIGER data, point sources from the EPA NPDES database, or locations of mine drainage from a variety of sources.)</p> <p>Estimates of percents are performed using standard EMAP methodologies.</p>

Section 2, EMAP Data

Table 2-5. Continued

MAIA–SURFACE WATERS	
Data Management, QA/QC, Standards, & Long-Term Maintenance	Data are maintained by WED in SAS and Arc/Info on a UNIX server that will be updated to NT. Codes for species and other parameters are unique EMAP codes to fit the SAS field limit of eight characters (for example, the first four letters of a genus and first four letters of a species are combined to make the code for a particular specimen). These codes are documented in the metadata. As data sets are used and updated, different versions of the data set may be kept by Indicator Leads and ORD-Corvallis, and may require version control.
Data Distribution	The watershed metrics will be distributed from the EMAP Public Web Site in a watershed characteristics database that contains stream ID’s matched to the metrics measured along that stretch of stream. Some of the data will be in SAS files, some will be in GIS (Arc/Info export) files. Other more detailed data can also be distributed in spreadsheets on request from WED. Raw data are maintained at WED. The volume and status of data sets are described in Appendix C, Inventory of EMAP Data.
Data Documentation	<p>Metadata for data sets are produced in EMAP Data Catalog format and targeted to users who have a good technical understanding of the content. Data sets are in the process of being documented.</p> <p>Descriptions of field and lab data collection and analysis methodology will be placed on the EMAP Public Web Site in .PDF (Adobe Acrobat) format. Descriptions of metrics calculations may also be created and placed on the web site.</p>

2.3.1.3 Mid-Atlantic Integrated Assessment–Landscape Ecology

Note: Most of the data integration, analysis, methodology, and other issues for this MAIA working group are the same as those for the Landscape Ecology Working Group, which is described in Section 2.3.3, Landscape Ecology, and contains a full description of Landscape Ecology data activities.

The purpose of data collection and analysis in the MAIA–Landscape Ecology group is to generate landscape indicators and use them to conduct regional ecological assessments. The focus is on integrating data from multiple scales—from site data to remote sensing images of entire regions—into the indicator development and assessment protocols. The result of the work is the MAIA landscape atlas (U.S. EPA 1998e), which presents the indicators, assessment protocols, and results of the integrated assessments.

Section 2, EMAP Data

Table 2-6. MAIA–Landscape Ecology Data Management

	MAIA–Landscape Ecology
Data Collection & Existing Sources	Please see Section 2.3.3, Landscape Ecology (below) for a description of data sources used by Landscape Ecology.
Data Aggregates & Products	The primary data product is the MAIA landscape atlas, which is posted on the EMAP Public Web Site (EMAP 1998). See Section 2.3.3, Landscape Ecology, for a full description of data products developed by this Working Group.
Georeferencing and GIS Products	The entire data set consists of geographically referenced data from the MRLC Interagency Consortium; see Section 2.3.3, Landscape Ecology (below), and the MRLC web site for more information on MRLC data.
Data Integration Issues	See Section 2.3.3, Landscape Ecology (below), for a description of data integration issues.
Methods, Algorithms, Models, Equations, & Indicators	See Section 2.3.3, Landscape Ecology (below), for a description of indicators developed by Landscape Ecology for use in determining the condition of landscapes, estuaries, and surface waters.
Data Management, QA/QC, Standards, & Long-Term Maintenance	See Section 2.3.3, Landscape Ecology (below), for a description of data management in this Working Group.
Data Distribution	Landscape indicators, assessment methodology, documentation, and the MAIA landscape atlas are available on the EMAP Public Web Site. The Atlas is also available on CD-ROM from the Landscape Ecology Working Group.
Data Documentation	See 2.3.3, Landscape Ecology (below), for more information on Landscape Ecology documentation.

2.3.2 Intensive/Index Sites

The purpose of the Intensive/Index Sites Working Group is “to develop a national framework for integration and coordination of environmental monitoring and related research through collaboration and building upon existing networks and programs and to support CENR goals for establishing long-term, intensive monitoring sites.” The program goal is to identify a set of national monitoring research sites to test intensive, multimedia, long-term monitoring at fixed index sites (i.e., intensive or index sites).

The Intensive/Index Sites Working Group is currently working with the CENR to identify the appropriate characteristics for site selection. It will demonstrate that sites chosen using these criteria can produce useful information for assessing ecological indicators and that selected common indicators can be used in this site format to provide information needed to forecast future risks to sustainability, particularly for stresses associated with global exposures.

Section 2, EMAP Data

The Intensive Sites research supports EMAP's efforts to address two major objectives in the current program that require sampling at fixed index sites repeatedly over a long period of time. The first objective is the need to incorporate existing long-term monitoring data into integrated assessments wherever possible. Historically, the majority of data collected by non-EMAP monitoring programs has been based at fixed stations. However, 1990–1995 EMAP data were collected under a probabilistic sampling plan. Although probabilistic sampling is becoming more prevalent among a number of non-EMAP research groups as EMAP has demonstrated the success of the approach to local and regional monitoring organizations, the majority of existing data that EMAP uses in its assessments will be from fixed sites. The problem of integrating data from fixed sites and random sampling sites is a non-trivial exercise for EMAP. In order for EMAP researchers to successfully conduct their assessments, EMAP will study the characteristics of intensive sites and refine data integration methodologies.

The second EMAP objective addressed by Intensive Sites research is to help EMAP researchers develop monitoring and assessment approaches that effectively characterize both the status of environmental resources and long-term trends in these systems. Although 1990–1995 data were useful for estimating natural resource status, they did not contribute to detecting long-term trends because the random sampling design resulted in pooling spatial and temporal variability. Monitoring at intensive sites can be used to detect long-term trends because it includes only temporal variability, thereby increasing the sensitivity to detect trends over time. The Intensive/Index Sites research program will employ fixed stations at which a consistent set of parameters are measured repeatedly over an extended period of time. These data will be used to refine the methodology for integrating data from related random sampling studies that have been conducted.

To accomplish the objectives outlined above, the Intensive/Index Sites Working Group is coordinating two primary efforts that are now in their early stages: the DISPro and the CISNet monitoring program. These efforts are part of a larger context of monitoring long-term fixed sites and therefore must coordinate with existing research programs in other agencies. The details of the studies and the cooperating agencies are summarized in Sections 2.3.2.1 and 2.3.2.2.

2.3.2.1 Demonstration Intensive Sites Project

DISPro is a joint effort between EMAP and the NPS to develop a demonstration of an intensive site network of monitoring and research locations throughout the United States utilizing the Nation's parklands and "outdoor laboratories." The NPS has an extensive program of air quality and environmental monitoring in the national parks; DISPro participates in monitoring a set of parameters at existing NPS sites in 12 national parks. EMAP and NPS have established 14 sites to demonstrate appropriate site selection, measurements, and analyses for assessing local, regional, and national ecological condition. These sites were chosen according to selection criteria cited in the EMAP Research Plan (U.S. EPA, 1997b) and include such factors as site accessibility, completeness and duration of monitoring records, and diversity of ecological communities. Information from these

Section 2, EMAP Data

sites will also help to identify research hypotheses, particularly those related to process-level phenomena, for further investigation (DISPro 1998a, 1998b).

The purpose of DISPro is to conduct long-term effects research for atmospheric data at sites in the national parks that already have 10–15 years of air quality data for comparison. (e.g., air deposition). DISPro focuses its studies on filling gaps in existing NPS monitoring. The intent of the program is to initiate a consistent monitoring program for the atmospheric parameters to be measured at each site; this initial monitoring will be followed up by monitoring of other media in order to examine the effects of environmental stressors of importance at each of the sites.

DISPro will interact with the ORD Regional Integrated Assessments studies wherever there are national parks in the Regional Assessment pilot study area. For MAIA, the Shenandoah National Park is in the study area. The next ORD Regional Assessment (the Western Pilot Study) will cover Regions VIII, IX, and X, and will include several national parks.

DISPro has initiated two new programs in cooperation with the NPS, including:

- ultraviolet-B (UV-B) monitoring; and
- individual air-quality research grants.

These programs are summarized below.

In order to conduct assessments of the effects of air quality on selected resources, DISPro is coordinating with and using data from a number of related NPS and EPA atmospheric and ecological monitoring programs, including:

- NPS routine air-quality monitoring;
- EPA Clean Air Status and Trends Network (CASTNet) National Dry Deposition Network (NDDN);
- National Atmospheric Deposition Program/National Trends Network (NADP/NTN) Wet Deposition Data;
- Interagency Monitoring of Protected Visual Environments Program (IMPROVE) visibility monitoring; and
- NPS NRI and Monitoring Programs.

These ongoing programs are briefly described at the end of this section and references are provided for additional information.

Section 2, EMAP Data

The two components of DISPro are UV-B monitoring and individual research grants. The UV-B monitoring has been initiated, but the individual research grants have not yet begun. More information for these will be added as it becomes available. Information about DISPro can be found on the EMAP Public Web Site.

UV-B Monitoring Program

UV-B monitoring is a cooperative project between EPA and NPS to fill a gap in the existing set of air-quality monitoring parameters.

The purpose of collecting UV-B data is to measure full-sky solar UV-B and some ultraviolet-A (UV-A) spectral flux, from which absolute irradiance and total column ozone concentrations are calculated.

Table 2-7. EMAP Intensive Sites—DISPro—UV-B Monitoring

EMAP Intensive Sites—DISPro—UV-B Monitoring	
Data Collection & Existing Sources	<p>Currently, the National Ultra-Violet Monitoring Center (NUVMC) at the University of Georgia (UGA) downloads UV-B measurements from Brewer high spectral resolution spectroradiometers at selected NPS and urban monitoring sites. The NUVMC is part of the UGA/EPA UV Monitoring Network (UVMN) that operates and maintains a group of high spectral resolution spectroradiometers throughout the United States.</p> <p>The Working Group will also use related air quality parameters and use the related historical NPS air monitoring data from these sites in future assessments.</p>
Data Aggregates & Products	Types of analyses and data products will be determined in future research. At present, summary plots of raw data are available at the NUVMC web site (NUVMC, 1998), and will soon be available in a standalone database on an EPA server.
Georeferencing and GIS Products	All data collection sites have a latitude and longitude in the database. Locations are reported in different formats by different researchers and would have to be reconciled for use in assessments. There has not been any specific coordination on this data collection issue among the different data collection groups.
Data Integration Issues	<p>The Working Group reported that the data are not currently available in formats that allow easy integration. Different researchers within the Working Group keep air measurements in different formats, so data sets are reconciled individually for aggregation or analysis.</p> <p>There is a need for standardization of data integration goals and techniques, tools, formats, and presentation. A consistent, long-range sampling plan needs to be documented for each site as well as for the overall program. No continuity of sampling locations is built into the long-term monitoring plan.</p>

Section 2, EMAP Data

Table 2-7. Continued

EMAP Intensive Sites—DISPro—UV-B Monitoring	
Methods, Algorithms, Models, Equations, & Indicators	Individual research and data repository sites use their own data analysis tools.
Data Management, QA/QC, Standards, & Long-Term Maintenance	NUVMC downloads data from the collection sites daily and conducts some preliminary processing, including a screen to catch immediate problems and errors in the data, and some validation and calculations. The result is a raw data file, as well as a data file that has been through first-level QA/QC. NUVMC also calculates plots of total UV (DUV/irradiance) and total column ozone. These plots are currently posted on the NUVMC web site (NUVMC 1998). Methodology and data reduction Standard Operating Procedures (SOP) will be written by the contractor.
Data Distribution	<p>The data are intended for dissemination to government and non-government scientists and interested parties (general public).</p> <p>UV-B data will be distributed to researchers from a standalone Oracle database on the EPA web server. This system will be password protected. This site is currently under development and will contain the raw data, geographic coordinates, and capabilities for conducting calculations (e.g., for plots). Summary data and metadata will be stored on the EPA Environmental Information Management System (EIMS) database and on a public web site (under development). The Regional Vulnerability Assessment Program (ReVA 1998c) web site will also point to the EIMS and standalone sites, and will store selected parameters and metadata. Currently, it appears that two calculated values (total column ozone and UV-B Index) will be stored at this site for access via password-protected accounts. These data and the associated metadata will become the authoritative site for UV-B researchers within EMAP (see Appendix B subsection B.5.1, Intensive/Index Sites Program Overview).</p> <p>A subset of the data that has been QA/QC'd by ReVA (e.g., total column ozone) will also be submitted to the Canadian WOUDC (WOUDC 1998) for long-term archival and distribution to secondary users, including the general public. Data at this site will be processed with the WOUDC validation and proofing techniques.</p>
Data Documentation	Metadata will be created by the field support contractor and will include site information (log sheets, audits, calibration records, maintenance records). The documentation will reside on the EIMS and standalone web sites. The NUVMC already has a minimal documentation set, but more detail will be added.

Individual Research Grants Program

This program will provide EMAP research funds to individual external researchers to conduct intensive site monitoring on a wide variety of topics to provide information about sampling design, indicator development, data assessment, and integration of random designs with index site designs. The following six projects have been selected, representing a variety of research areas:

Section 2, EMAP Data

- **Michigan Technological University**—“Below ground ecosystem function: Merging long-term climate monitoring with soil, root, and foodweb dynamics to understand mechanisms regulating C and N transformations.” Denali National Park;
- **Boyce Thompson Institute of Plant Research**—“Using the inter-relationships of stable isotopes in natural abundance as indicators of environmental stress and ecosystem vitality” Big Bend National Park, Glacier National Park and Sequoia - Kings Canyon National Park;
- **Institute of Ecosystem Studies**—“Atmospheric deposition in mountainous terrain: Scaling up to the landscape.” Acadia and Great Smoky Mountain National Parks;
- **University of Utah**—“Nitrogen deposition and UV-B stressor impacts in Canyonlands National Park as affected by climatic pulse events”;
- **University of Maine**—“Inferring regional patterns and responses in N and Hg biogeochemistry using two sets of gauged paired-watersheds.” Acadia National Park; and
- **U.S. Forest Service, Pacific Southwest Research Station**—“Does N deposition mitigate ozone injury to ponderosa pine?” Sequoia - Kings Canyon National Park.

These projects involve 8 of the 14 DISPro national parks and are intended to demonstrate the utility of index sites. They will not generate a large amount of data; any data that may be collected can be handled by EMAP-IM (AED) and made available via the EMAP-IM system. The grants will begin in 1998, and more information about the projects will be provided as it becomes available.

In order to plan research programs and obtain data for its assessments, DISPro coordinates with existing air quality monitoring programs in EPA and NPS. EMAP does not collect or manage data in these programs, but uses the data. Some of these programs are briefly described below.

National Park Service Routine Base Air Quality Monitoring Program

The NPS conducts ongoing air-quality monitoring at long-term stations in the national parks and collects such parameters as NOX and SOX. DISPro/EMAP will coordinate closely with these data efforts by collecting the UV-B data at a subset of these routine monitoring stations and using data from the monitoring records at these sites. NPS submits all of the ozone, sulfur dioxide, and meteorological data to the EPA Aerometric Information Retrieval System (AIRS; a computer-based repository of information about airborne pollution administered by the EPA Office of Air Quality Planning and Standards, which is part of the EPA Office of Air & Radiation) database, which makes the annual data summaries available on the AIRS web site (AIRS 1998). The AIRS web site is currently fully accessible only to users with an account, but accessibility is being improved. However, yearly averages are insufficient for most research purposes. Researchers cannot easily access the hourly data needed for response models and statistical analyses; access to these data on the AIRS web site will be implemented soon. The 1-hourly data that NPS submits to AIRS will be

Section 2, EMAP Data

made available at the AIRS site within a year. Electronic files of NPS data can be obtained directly from NPS by request. Query access to the NPS Oracle database will soon be available.

EPA Clean Air Status and Trends Network/National Dry Deposition Network

EPA CASTNet (CASTNet 1998) was established in 1987 and comprises a network of monitoring stations across the U.S. to monitor the results of emission reductions. The majority of the monitoring stations are operated under contract to EPA's Office of Air and Radiation. Of 71 total stations, 67 measure dry deposition, 18 measure wet deposition, 48 measure ozone, and 9 measure visibility. Rates are calculated using atmospheric concentrations, meteorological data, and information on land use, vegetation, and surface conditions. Of the 67 monitoring stations measuring dry deposition, 48 form the core of the network and were formerly known as the NDDN (NDDN 1998). The other 19 stations are located in national parks or other Class-I areas, which are areas designated by Congress as deserving special protection against air pollution (these stations are mostly located in the Western U.S. at locations where the NPS has been measuring ozone for several years).

CASTNet's ozone, sulfur dioxide, and meteorological data are submitted to the EPA AIRS database. AIRS is fully accessible only to people with an account at the AIRS Home Page (AIRS 1998). The filter-pack data are partially available at the CASTNet web site. Currently, only dry deposition data are available online, but other data will be available soon. Wet deposition data may be posted on the CASTNet web site in the future or CASTNet will provide a link to other sites where it is stored. Three of the six years of wet deposition data are only in the AIRS database; the other three years are on the web site (NPS 1998b). CD-ROMs of the entire CASTNet database may be obtained upon request.

National Atmospheric Deposition Program/National Trends Network Wet Deposition Data

The NADP/NTN (NADP/NTN 1998) collects wet deposition data at a nationwide network of more than 200 precipitation monitoring sites in the continental United States, Alaska, and Puerto Rico. The network is a cooperative effort between many different groups, including the State Agricultural Experiment Stations, USGS, USDA, and numerous other governmental and private entities. The purpose of the network is to collect weekly data on the chemistry of precipitation in order to monitor geographical and temporal long-term trends. Data are analyzed for hydrogen (acidity as pH), sulfate, nitrate, ammonia, chloride, mercury, and base cations (e.g., calcium, magnesium, potassium, and sodium). This effort is considered the Nation's primary source for wet deposition data.

The NADP has two subnetworks: the MDN (MDN 1998), which has collected weekly precipitation samples at more than 20 sites to track mercury deposition in lakes and streams on a regional basis since 1995; and the Atmospheric Integrated Research Monitoring Network (AIRMoN) (AIRMoN 1998), which collects daily samples from a network of nine sites and analyzes samples for the same constituents as the NADP/NTN samples to study precipitation chemistry trends with greater temporal resolution.

Interagency Monitoring of Protected Visual Environments Program

The IMPROVE program is a joint effort by the NPS, USFS, USFWS, Bureau of Land Management (BLM), and EPA to study atmospheric visibility in the national parks. The prime contractor is the University of California (Davis), which makes data collected under this project available via the Internet on an anonymous FTP site (IMPROVE 1998) which has ASCII files of data by individual site location. Some data are not currently on the FTP site and must be requested directly from the contractor (e.g., optical measurements with nephelometers and transmissometers and other equipment).

National Park Service Natural Resource Inventory Monitoring in National Parks

This program, not yet fully funded, will include biological survey data (baseline inventories and long-term monitoring data) collected in the parks (e.g., amphibians). These efforts will be coordinated with other agencies, and resources will be shared as appropriate. For more information on program scope, see the National Park Service monitoring web site (NPS 1998d).

2.3.2.2 Coastal Intensive Sites Network (CISNet)

The coastal component of the Intensive/Index Sites Working Group is the CISNet, which will establish pilot sites for the development of a network of intensive, long-term monitoring and research sites around the U.S. marine and Great Lakes coasts. This project represents an inter-agency effort between EPA/ORD, NOAA, and the National Aeronautics and Space Administration (NASA). EPA and NOAA will fund individual research grants to conduct field research and monitoring studies to develop ecological indicators and investigate the ecological effects of environmental stressors. NASA will cooperatively fund studies to develop a remote sensing capability to complement the field studies. Information about the types of studies that will be funded under CISNet can be found on the CISNet web site (CISNet 1998).

CISNet's objectives are to:

- develop a sound scientific basis for understanding ecological responses to anthropogenic stresses in coastal environments, including the interaction of exposure, environment/climate, and biological/ecological factors in the response, and the spatial and temporal nature of these interactions;
- demonstrate the usefulness of a set of intensively monitored sites for examining short-term variability in long-term trend behavior in the relationships between changes in environmental stressors, including anthropogenic and natural stresses, and ecological response; and
- provide intensively monitored sites for developing and evaluating indicators of change in coastal systems.

Section 2, EMAP Data

Table 2-8. Intensive Sites—Coastal Intensive Sites Monitoring Data Management

	Intensive Sites—Coastal Intensive Sites Monitoring
Data Collection & Existing Sources	Monitoring data will be collected at up to 41 pilot sites chosen by EPA as part of a network of intensive, long-term monitoring and research sites around the U.S. marine and Great Lakes coasts. Data sources, volumes, and other details will not be known until specific projects have been selected for funding. Areas of study may include: <ul style="list-style-type: none"> • development of indicators of coastal ecosystem integrity and sustainability; • assessment of temporal and spatial variability problems in environmental measurements; • effects of nitrogen and phosphorus effects on coastal systems; • effects of stressors on coastal systems, examination and evaluation of the effects of anthropogenic stressors on coastal systems; and • development of remote sensing capability.
Data Aggregates & Products	No projects begun yet.
Georeferencing and GIS Products	No projects begun yet.
Data Integration Issues	No projects begun yet.
Methods, Algorithms, Models, Equations, & Indicators	No projects begun yet.
Data Distribution	No projects begun yet.
Data Management, QA/QC, Standards, & Long-Term Maintenance	No projects begun yet.
Data Documentation	No projects begun yet.

2.3.3 Landscape Ecology

The mission of the Landscape Ecology Working Group is to “...initiate research in landscape ecology in order to incorporate meaningful indicators of land-cover configuration into regional ecological assessments.” (U. S. EPA 1994c, 1995d). Research will focus on integrating data from multiple scales (from field monitoring data at individual sites to remote sensing images of entire regions) to develop landscape indicators and assessment protocols and conduct landscape assessment studies. Some of the indicator development and integrated assessments will focus on estimating landscape change over a 20-year period.

Section 2, EMAP Data

Landscape studies are being conducted for both EMAP and ReVA. For EMAP, Landscape Ecology is developing landscape indicators and assessments of status and trends for selected resources (e.g., water quality, habitat quality). For ReVA, they are using the indicators to evaluate or assess resources for their potential vulnerability to future degradation as a result of multiple stressors. There is a synergy between EMAP and ReVA tasks, since some of the ReVA vulnerability assessments rely on indicators and data developed for the EMAP status and trends program. ReVA and EMAP need to work together to make data accessible to both groups.

Landscape Ecology is conducting two main projects for EMAP:

- MAIA landscape atlas (U.S. EPA 1998e) (see Section 2.3.1.3, MAIA Landscape Ecology); and
- Support of R-EMAP Projects (in cooperation with EPA Regions)
 - ❑ Region IV—Savannah River Landscape Analysis
 - ❑ Region VII—Landscape Analysis and Characterization to Support Regional Environmental Assessment Project
 - ❑ Region VIII—Integration of Upland and Riparian Stream Condition Monitoring for Intermediately Sized Watersheds on Rangelands
 - ❑ Region IX— Bioassessment of Water Quality in the Humboldt River, Nevada.

Related projects for ReVA are listed in Appendix B, Data Management Needs and Practices of EMAP Working Groups.

Section 2, EMAP Data

Table 2-9. Landscape Ecology Data Management

	Landscape Ecology
Data Collection & Existing Sources	Most of the data used in this program consists of existing satellite imagery. Data sources for landscape ecology includes images from the MRLC (MRLC 1998a, 1998b) databases (1992–1993), although Advanced Very High Resolution Radar (AVHRR) and EPA North American Land Characterization (NALC) satellite images are also be used for lower resolution analyses. Landscape Ecology needs to access data from EMAP Resource Groups (especially Surface Waters, Forests, and Agroecosystems) and current Working Groups (especially R-EMAP and MAIA-Surface Waters). The EPA RF3 database is a key basemap for their efforts; the quality and completeness of this database varies widely by region, depending on whether it has been updated. The Region III data have been updated, but corrected data are not available for the whole MAIA area (some of which falls in EPA Region II and Region IV).
Data Aggregates & Products	Data aggregates are produced for publication in atlases of environmental condition for the studied regions and include ecological indicators of landscape condition, as well as assessments of landscape condition using those indicators. To date, aggregates have been produced for the MAIA landscape atlas (U.S. EPA 1998e).
Georeferencing and GIS Products	The entire data set consists of geographically referenced data from the MRLC Interagency Consortium (MRLC 1998a). MRLC is a national land cover database comprising multiple data layers, including: <ul style="list-style-type: none"> • general land cover based on a modified NOAA Coastal Change Analysis Program (C-CAP) legend; • source materials used in the classification (including the georeferenced Landsat TM data, ancillary data, and digital validation data); • disaggregated, stratified spectral classes (or their equivalent) used to produce the general classification; and • detailed land cover from the USFWS Gap Analysis Program (GAP), and where available, from C-CAP, USGS NAWQA, and USFS.
Data Integration Issues	Landscape Ecology integrates the geographic data sets with field sampling data sets and will reconcile locations, resolution, etc. The challenge of this work is to reconcile data collected and stored with different spatial scales, standards, and methods of measurement. For some data available from NFS and USDA the locational data are purposefully dithered to prevent use in regulatory contexts and to ensure unbiased sampling. Data integration issues also involve inadequate access to existing EMAP data (e.g., geographic locations for sampling data). EMAP-IM (AED) is working with Landscape Ecology to obtain needed data sources from other groups.

Section 2, EMAP Data

Table 2-9. Continued

	Landscape Ecology
Methods, Algorithms, Models, Equations, & Indicators	Primary tools produced for MAIA are the indicators of ecological condition. For other projects, Landscape Ecology also produces methodology for assessing the condition of natural resources based on landscape scale monitoring and comparing their landscape indicators with ecological indicators from other Working Groups (e.g., Forest Stands vs. Chlorophyll Index).
Data Management, QA/QC, Standards, & Long-Term Maintenance	All raw data (e.g., images, field data) and data in progress are managed and maintained by the Landscape Ecology Working Group. Some of the data are managed at ORD-Landscape Ecology Branch (Environmental Sciences Division, NERL) and a USGS office at the TVA. Summary data (indicators, assessment results) are archived on the EMAP Public Web Site in PDF and Arc/Info export format, respectively.
Data Distribution	Data distributed landscape atlas products (e.g., Chesapeake Bay, Mid-Atlantic), as well as landscape analysis methodology that other groups may apply to their data. Data distribution issues for Landscape Ecology include the need to make large GIS files available to research partners as well as to the public. See Section 3, Information Management Needs and Requirements, for a discussion of the issue of providing access to these data for different kinds of users.
Data Documentation	Landscape Ecology considers it a priority to produce documentation that is distributed with each data product because there are many ways to do the same analyses (e.g., Normalized Difference vegetation Index [NDVI]). Therefore it is critical for legal and scientific purposes to have good descriptions of data quality and analytical procedures. The Landscape Ecology group receives many requests for data and interpretations of the data. These requests come from a variety of users with many different objectives, such as scientists, the public, and researchers. Results of these assessments may be used to support monitoring plans or court litigation. Therefore, the methodology used to analyze and interpret the data are of critical importance to future users. All observations made using these data depend on interpretations of digital data which always contain some assumptions. The documentation must be complete and explicit for the data to be interpreted in context. Existing documentation fulfills minimum requirements outlined in the 1995 Mid-Atlantic Landscape Ecology workplan. The Working Group will convert the documentation to be compatible with the FGDC spatial metadata standards. The Working Group needs additional resources to complete Data Directory entries and FGDC documentation (assistance may be provided by ReVA). Landscape Ecology has requested technical guidance to support their efforts to achieve FGDC compliance; EMAP-IM (AED) can provide examples of metadata templates and tools from EMAP, the NOAA Coastal Services Center (CSC), and other sources.

2.3.4 Regional EMAP

The goal of the R-EMAP Working Group is “to assist in incorporating the latest science on ecological monitoring into EPA Regional, state, tribal, and local decision-making processes, and to reduce the scientific uncertainty in local decisions using environmental risk-based management.” (U.S. EPA 1997b). R-EMAP projects are coordinated by EPA Regional Offices and conducted by a combination of EPA, state, local, academic, and other researchers.

R-EMAP projects focus on monitoring issues of concern to the regional scientists, therefore, the topics, data collection, products, management, and distribution of data vary by region. More detail about how these activities are accomplished in several Regions can be found in Appendix B, Data Management Needs and Practices of EMAP Working Groups.

R-EMAP data focuses on particular issues within the Regions that are likely to be critical sources of data for ORD Regional Assessments such as MAIA. For example, R-EMAP Region III’s study of the state of streams in the Mid-Atlantic Highlands coordinates their data collection and analysis efforts with MAIA-Surface Waters.

Table 2-10 lists the projects that have been completed or are ongoing in Round 1 (1993) and Round 2 (1995) of R-EMAP. Proposed future projects now being reviewed for the next round of funding are not included in this list, but will be added as they are approved.

Table 2-10. R-EMAP Projects by Region

Region I	Fish tissue contamination in Maine. Mercury deposition and atmospheric concentrations in New England. Assessment of mercury in hypolimnetic lake-bed sediments of Vermont and New Hampshire.
Region II	Baseline study of New York/New Jersey Harbor sediments. Continuation of the Harbor study in a trend assessment of sediment quality and development of indicators for the New York/New Jersey Harbor (basis for New York/New Jersey Harbor Estuary Program’s long-term monitoring program for the Harbor).
Region III	State of streams in the Mid-Atlantic Highlands including all geographic areas of Region III except the coastal plain and the piedmont.
Region IV	Everglades ecosystem assessment (system-wide research and monitoring study conducted of mercury contamination, eutrophication, habitat alteration, and hydropattern modification issues). Savannah River project.

Section 2, EMAP Data

Table 2-10. Continued

Region V	<p>Corn Belt project.</p> <p>Develop biological indicators for watersheds and assess the status of wadeable streams to understand the spatial evolution of northern lakes and forests.</p> <p>Condition of the St. Louis River embayment of Lake Superior.</p>
Region VI	<p>Toxic substances characterizations for selected Texas estuaries (Galveston Bay, Corpus Christi Bay)—follow-up to EMAP Estuaries (Louisianan Province) monitoring program.</p> <p>Application of a probabilistic approach to determine the extent and effects of stream habitat degradation and fish community integrity in eastern Texas streams.</p>
Region VII	<p>Status of stream water quality of Nebraska, Kansas, and Missouri and develop an Index of Biological Indicators to measure the health of fish and habitats.</p> <p>Landscape analysis and characterization to support regional environmental assessment of status of stream water quality in Kansas, Nebraska, and Missouri.</p> <p>Resampling Nebraska streams from Round 1 study and assisting the state of Nebraska with using probabilistic design and EMAP indicators for rotating basins studies.</p>
Region VIII	<p>Assessment of metals impact in headwaters streams within mineralized areas of the Southern Rockies ecoregion and development of a Regional Biotic Index.</p> <p>Grazing impacts on rangeland conditions in Utah with Utah State University and ORD Characterization Research Division.</p>
Region IX	<p>Assessment of the Southern California Bight from Pt. Conception to the Mexico border.</p> <p>Surface water assessment in natural streams and constructed conveyances of California's Central Valley (using EMAP sampling methodology).</p> <p>Dewatering issues in aquatic systems of the Humboldt River watershed, Nevada (Basin and Range Province) and development of new EMAP protocols for arid stream assessments.</p>
Region X	<p>Ecological assessment of streams and riparian areas in the Upper Deschutes (Oregon) and Upper Chehalis (Washington) basins.</p>

R-EMAP Data Collection and Management

Data collection activities of R-EMAP are briefly summarized below; for more detail, see Appendix B, Data Management Needs and Practices of EMAP Working Groups, in which data activities of selected Regions are described.

R-EMAP projects collect monitoring data according to EMAP protocols (sampling, analysis, etc.) to solve problems of significance to each EPA Region.

Section 2, EMAP Data

Table 2-11. R-EMAP Overall Program Data Management

	R-EMAP
Data Collection & Existing Sources	Each project collects new data and uses existing data from a number of widely dispersed data sources, including states, regional boards, EPA Regions, and EMAP. R-EMAP data cover a wide variety of data types and individual projects can cover multiple media (e.g., mercury in air, water, sediment, and fish tissue).
Data Aggregates & Products	The types of data products of R-EMAP data vary by region; see Appendix B (Data Management Needs and Practices of EMAP Working Groups) and Appendix C (Inventory of EMAP Data), for more detail.
Georeferencing and GIS Products	A variety of georeferenced sampling data sets and GIS coverages are being produced in the Regions, in addition to Landscape Ecology integrated assessments that are being prepared.
Data Integration Issues	Data are collected at a wide variety of both spatial and temporal scales, and integration efforts involve data from local scales (heavily sampled streams) to landscape ecology GIS coverages (3-state regions). Data integration may introduce bias due to differences in collection methodology, sampling scale, and analytical approaches.
Methods, Algorithms, Models, Equations, & Indicators	R-EMAP projects are primary users of methods developed by other EMAP programs (e.g., MAIA Habitat Quality Index) but such information is currently difficult to locate.
Data Management, QA/QC, Standards, & Long-Term Maintenance	Data are managed by the researchers who collected them, including university laboratories, state and federal agencies, etc. These groups use a wide variety of data management tools, formats, and standards. No standards were available for the existing projects, but EMAP has provided standards to ORD's MED for future projects. ORD's WED has collected and is analyzing and managing some of these data sets in SAS. WED will provide the data to the EPA Regions and EMAP-IM (AED) in ASCII or SAS or spreadsheets as appropriate. See Appendix B, Data Management Needs and Practices of EMAP Working Groups, for more detail on data management in selected R-EMAP groups.
Data Distribution	Many R-EMAP groups plan to enter their data into non-EMAP data repositories, such as modernized STORET, EPA AIRS, and specialized databases like the MDN. See Appendix B (Data Management Needs and Practices of EMAP Working Groups) and Appendix C (Inventory of EMAP Data) for more information on repositories and distribution for R-EMAP data.
Data Documentation	Metadata standards were given to the Regions after the first round of projects was begun in 1993, but very few of the R-EMAP groups have developed the required documentation. Many of the data sets still reside with the project managers at universities and other sites with little or no documentation. EMAP-IM (AED) provides assistance to the EPA Regions or the R-EMAP programs for collecting and documenting the data. Ongoing projects (1995 to present) have been given EMAP standards for data organization and documentation, but the researchers require support and guidance from EMAP-IM (AED) to complete the requirements.

2.3.5 Ecological Indicator Development

The mission of the Ecological Indicator Development Working Group is to oversee the application of a standard approach and review process to the development and evaluation of indicators (measurable or calculated parameters). The goal of this standard approach is to ensure the propagation of robust, peer-reviewed, widely applicable diagnostic tools that can help define the status of critical environmental resources. The principal activity of the Working Group in carrying out its mission is to create and maintain a framework for evaluating and documenting indicators. This process consists of an evaluation of the indicator according to *Ecological Indicator Evaluation Guidelines* (U.S. EPA, in prep.) that are now being developed, and a formal peer review by a committee established by the Working Group that evaluates the quality and applicability of an indicator. Data generated by this working group will consist of documentation of the indicator and its evaluation according to the Guidelines and the peer review process. The role of the documentation is summarized in the EMAP Research Strategy: “Documentation of ORD evaluations will ultimately generate a dynamic and iterative base of knowledge on the strengths and weaknesses of individual indicators.” (U.S. EPA 1997a).

Indicators are being developed by many ongoing programs in ORD and other organizations. The effort to evaluate indicators is an ORD-wide initiative, and the role of the Working Group is to coordinate research and oversee the evaluation of indicators being developed by other Working Groups such as Landscape Ecology, MAIA–Surface Waters, and MAIA–Estuaries.

The Working Group’s goals are to:

- identify indicator priorities and use them to provide direction to ORD indicator research and development;
- provide a means to evaluate indicators for monitoring and assessment activities through the *Ecological Indicator Evaluation Guidelines* and a peer review process refereed by the Ecological Indicators Working Group;
- integrate research efforts performed intramurally, extramurally, and by other agencies and programs; and
- ensure that ORD research is responsive to the needs of clients (e.g., Program Offices) and users (e.g., risk assessors) by establishing and maintaining interactions with EPA Regions and program offices to remain responsive to risk assessors.

Initially, research will focus on the development and characterization of indicators that emphasize ecological components and functions to represent or reflect specific, well-known environmental values. This focus on specific environmental values will provide direct linkage to EPA’s existing environmental risk assessment process, which incorporates an analysis of environmental values in

Section 2, EMAP Data

developing risk assessment endpoints. The research will also strive to anticipate management needs for indicators of ecosystem integrity and sustainability.

The *Guidelines* will set the standards for what is expected of an indicator by applying 15 criteria to the indicator in four phases (U.S. EPA 1997b):

- Ecological relevance, relationship to environmental value to society;
- Estimation of logistics to apply the indicator (resources, time, equipment, etc.);
- Variability (seasonally, during index period, etc.); and
- Difference detection, indicator applications.

Each indicator must pass sequentially through each phase; no indicator can be evaluated under a later phase until it has passed through the previous phase.

Any indicator that meets the *Guidelines* will be evaluated by the Working Group for admission into the Peer Review process and, if it passes, the Working Group will set up an indicator evaluation panel and begin the review. The peer review panel will consist of scientific experts and risk assessors who can determine the strengths and weaknesses of the indicator. The peer review process is designed to:

- document evaluation of the indicator in an established sequence of steps (which could be used by any organization including EPA, state agencies, etc.);
- allow future users to understand the validity, utility, and requirements for implementing the indicator and the data that supports it;
- establish directions for future research into the areas of indicator strengths and weaknesses;
- allow an iterative process so that review steps will be repeated as indicators are updated in response to reviews; and
- evaluate all documents and other products produced by the Working Group.

The *Guidelines* and further information about the Ecological Indicators research will be made available on the EMAP Public Web Site in the future.

The principal information products of this Working Group in carrying out its mission will be:

- the Ecological Indicator Evaluation Guidelines, which specify the methodology for evaluating indicators;

Section 2, EMAP Data

- indicator documentation that was prepared to present the indicator to the peer review panel (including methodology, references to supporting data, etc.); and
- documentation of peer reviews (including strengths and weaknesses of the indicator, appropriate uses, reviewer concerns, etc.)

Table 2-12. Ecological Indicator Development Data Management

	Ecological Indicator Development
Data Collection & Existing Sources	<p>Researchers will collect or use monitoring data on an as-needed basis to conduct pilot studies to test the indicator or introduce data to support the indicator presentation to the peer review panel. These data can include new monitoring data or existing data sources. In either case, researchers are responsible for data quality and QA/QC. If researchers collect new monitoring data, they will conduct their own data management and analysis (the Working Group may need to develop a management plan or standards for handling these data). Existing data sources will be located and obtained by researchers and are expected to come from manuscripts, field notes, published literature, historic databases, and new pilot sampling programs. Data sources used will cover a variety of disciplines, since indicators are being developed in a number of selected areas of importance to EMAP, including:</p> <ul style="list-style-type: none"> • forests; • fresh waters (lakes, streams, wetlands); • estuaries and coastal wetlands; • landscapes (across resource types); and • integration of whole ecosystem. <p>The data sources collected or used will be included in the EMAP-IM system by being cited in the indicator presentation and peer review summary. Links to critical data sets will be made through the EMAP Data Directory.</p>
Data Aggregates & Products	<p>The main data product will be the Guidelines and the indicator evaluations. Development of the indicator will include data aggregates that will be managed and presented by the principal investigators developing the indices (not by the Ecological Indicators Working Group).</p>
Georeferencing and GIS Products	<p>N/A (will only apply to data used to test the indicator)</p>
Data Integration Issues	<p>N/A (will be the responsibility of individual Principal Investigators)</p>
Methods, Algorithms, Models, Equations, & Indicators	<p>The indicator information will be stored as text documentation, which will include description of the methodology, uses, strengths and weaknesses.</p>

Section 2, EMAP Data

Table 2-12. Continued

	Ecological Indicator Development
Data Management, QA/QC, Standards, & Long-Term Maintenance	<p>Researchers will conduct indicator QA/QC according to the Guidelines.</p> <p>Indicator evaluations can be stored on the EMAP Public Web Site, including responses to the Guidelines, indicator descriptions, peer review summaries, and links to or citations of supporting data.</p>
Data Distribution	<p>The indicator documentation and the Guidelines will be made available so that future users can obtain them and understand their use. The Working Group will concentrate its efforts on distributing and tracking these products, and EMAP-IM (AED) will assist with these tasks.</p> <p>The monitoring data used to test indicators will remain with the Principal Investigators. There are no plans to distribute the data sets involved in the indicator development via the EMAP Public Web Site. However, the data could be cited and documented to support the indicator evaluation. Direct exchange of these data among indicator researchers will be infrequent and will take place by direct contact between the individuals, so the Working Group does not see a need to index data sources in the EMAP-IM system or provide a web site for data exchange.</p>
Data Documentation	<p>The documentation of the indicator evaluation and peer review processes will consist of the following text-based documentation:</p> <ul style="list-style-type: none"> • The text of the Ecological Indicator Evaluation Guidelines; • Presentation of each indicator, including a description of the indicator and the responses to the Guidelines; • Peer reviewer summaries of the strengths and weaknesses of the indicator (based on the purpose for which it was designed); and • Links and references to any data sources used to demonstrate and justify indicators under the Guidelines.

2.3.5.1 Aquatic Mortality Monitoring Database

Beginning in 1998, EMAP will work with States develop a database that incorporates information about the mortality of marine and estuarine aquatic organisms on the Atlantic and Gulf coasts. Mortality events are important not only because of the loss of the affected organisms, but also because they may signal the presence of public health dangers or degraded environmental conditions. Knowing the nature, extent, and probable cause can ultimately lead to actions that minimize impacts and reduce the risk of recurrence. From an EMAP monitoring and information management perspective, consistent investigation and documentation of mortality events and epizootics can lead to a better understanding of changing environmental conditions at different spatial scales.

This database is modeled on the existing Gulf of Mexico Aquatic Mortality Network (GMNET) (coordinated by EPA’s Gulf Ecology Division through the Gulf of Mexico Program). GMNET includes mortality response teams from all five Gulf Coastal States and three federal agencies (EPA, USGS–BRD, and NOAA). Members of this intergovernmental network share the common goals, which are to:

Section 2, EMAP Data

- improve interstate communication among mortality response teams to improve the utility of the early warning system and raise the quality of response information;
- develop a network of scientists to provide chemical and pathological expertise to support efforts to determine the cause of mortality events; and
- provide place and time analyses of aquatic mortalities in the Gulf of Mexico so that the data can be related to other important events (hypoxia, red tide, El Nino, etc.) and can cumulatively serve as an indicator of ecological condition in the Gulf (GMNET 1998).

EMAP–IM (AED)’s focus will be to extend the database efforts to the Atlantic states and develop a comprehensive database that can be used by all participants; most coastal states support mortality response teams that investigate and determine probable causes of mortality events. Later, this work may be expanded to the Pacific coast.

This effort will involve:

- establishing communication among the states;
- merging and integrating mortality information to meet common goals while maintaining the identity and purpose of their individual state mandates (including adopting standard response approaches and techniques for investigating and documenting mortality events, and holding interstate training exercises to reinforce their use);
- collecting the same information and documenting all mortality events using the same database format and spreadsheet;
- merging data from all five states into a regional database that can be incorporated into a GIS presentation and analysis;
- using the database to demonstrate regional trends, areas of high and low activity, seasonal trends, or identify causes of mortality;
- characterizing relative conditions in the Gulf of Mexico over time and serving as a warning if conditions start to deteriorate rapidly;
- using the data to develop an “epidemiological” (or epizootiological) approach for understanding the environmental conditions that lead to disease and mortality;
- establishing direct cooperation among state response teams and scientists in a variety of disciplines to develop the most credible and scientifically-defendable diagnostic information; and
- extending information to the public on the Gulf of Mexico Program web site (GMNET 1998).

The desired results will be:

- high-quality regional data;
- improved reporting of events by fishermen, beach-goers, boaters, and residents;
- consistent and comprehensive coverage and reporting of events;
- useful information on relationships of mortalities to regional and climatic events such as red tides, El Nino, and global climate change;
- consistent and high-quality response and reporting efforts;
- regional information and a regional perspective, created by integrating data that are collected at state and local levels; and
- consistent documentation of mortality events, ultimately leading to the development of early-warning, status, and condition indicators that can support efforts to maintain the Gulf of Mexico as a productive habitat for living resources.

2.4 Conclusions

EMAP is producing a heterogeneous and geographically distributed set of data and documentation that are being made available to users through the EMAP-IM system and a number of non-EMAP data repositories. The missions of the Working Groups are diverse and dynamic, and will evolve over time to produce a variety of data types for widely distributed users who want access to the data to support new research and analysis. To support distribution of this rich set of information, EMAP-IM system components and standards must be open and flexible enough to adapt to changing needs and operate effectively within the diversity of information systems that hold EMAP data. The requirements for such a system are described in the following sections.

Section 3

Information Management Needs and Requirements

- 3.1 Introduction**
- 3.2 User Needs**
 - 3.2.1 General Users**
 - 3.2.2 Primary Users**
- 3.3 Recommended Guidelines for EMAP Data Sources**
 - 3.3.1 Types of EMAP Data Sources**
 - 3.3.1.1 ORD Data Sources**
 - 3.3.1.2 Non-ORD EMAP Data Sources**
 - 3.3.2 Recommended Guidelines for EMAP Data Management and Delivery**
 - 3.3.2.1 Data Collection**
 - 3.3.2.2 Data Aggregates**
 - 3.3.2.3 Data Integration**
 - 3.3.2.4 Data Delivery–Exchange and Distribution**
 - 3.3.2.5 Documentation**
 - 3.3.2.6 Data Archiving**
 - 3.3.2.7 Data Storage**
 - 3.3.2.8 Version Control**
- 3.4 EMAP–IM Functional Requirements**
 - 3.4.1 Track EMAP and Non–EMAP Data Relevant to EMAP Research**
 - 3.4.1.1 Provide Access to Individual EMAP Data Products by Tracking Them at the Widely Dispersed Data Distribution Sites**
 - 3.4.1.2 Provide Documentation of the Quality of External Data Sources and Metadata**
 - 3.4.1.3 Organize Data and Metadata for Ease of Retrieval and Updating**
 - 3.4.1.4 Keep Data Directory Entries Up-to-Date and Synchronized with Evolving Versions of EMAP Data and Data Sources Frequently Accessed by EMAP Users**
 - 3.4.2 Facilitate Rapid, Ad Hoc Data Exchange Among EMAP Researchers**
 - 3.4.3 Provide Tools, Standards, and Support to Users and Data Collectors**

- 3.4.3.1 **Maintain and Disseminate Standards for Data Collection, Management, Documentation, and Distribution**
 - 3.4.3.2 **Support EMAP Data Collectors and Users with Data and Metadata Preparation**
 - 3.4.3.3 **Participate in CENR Standards Development and Implementation**
 - 3.4.3.4 **Distribute EMAP Tools and Automated Procedures for Research Planning and Implementation, and Monitoring Network Design**
 - 3.4.4 **Maintain and Update EMAP–IM System (Components and Network Connections)**
 - 3.4.5 **Deliver EMAP Data and Information Managed by EMAP–IM (AED)**
 - 3.4.5.1 **Improve Access to EMAP Resource Groups Data Sets and Metadata**
 - 3.4.5.2 **Capture “Orphan” Data Sets in the EMAP–IM System**
- 3.5 **System Requirements**
 - 3.5.1 **Overall System Design**
 - 3.5.2 **System Components**
 - 3.5.2.1 **EMAP Data Directory**
 - 3.5.2.2 **EMAP Data Catalog**
 - 3.5.2.3 **EMAP Public Web Site**
 - 3.5.2.4 **EMAP Internal Web Site**
 - 3.5.3 **System Configuration—Software, Hardware, Network, and Online Resources**
 - 3.5.3.1 **Software and Data Processing Resources**
 - 3.5.3.2 **Hardware Resources**
 - 3.5.3.3 **Network Availability and Capacity**
 - 3.5.3.4 **Storage and Distribution**
 - 3.5.3.5 **Data Security**
- 3.6 **Conclusions**

Section 3, Information Management Needs and Requirements

The primary purpose of the EMAP–IM system is to deliver relevant scientific data and information to EMAP researchers and other users in a timely, user friendly fashion. EMAP–IM and Working Group research partners share this responsibility. Working Groups manage and deliver the data to EMAP and other data repositories. EMAP–IM supports the Working Groups by providing data submission standards, metadata entry tools, funding, assistance with completion of Data Directory entries, and maintaining the EMAP–IM infrastructure. For this shared stewardship to succeed, certain requirements for EMAP–IM (AED) and Working Group activities and system features that link these efforts together are outlined in this section.

3.1 Introduction

The requirements for the EMAP–IM system can be divided into four categories, which are discussed in separate subsections below:

- **User Needs**, which represent the broad categories of EMAP–IM users and their needs for information and functionality;
- **Data Source Requirements**, which represent the responsibilities of EMAP researchers and other data sources for making data and documentation available to EMAP users;
- **Functional Requirements**, which represent the range of capabilities the EMAP–IM system must be able to accomplish and the support EMAP–IM (AED) must provide in order to ensure long-term data availability; and
- **System Requirements**, which describe the system configuration, hardware, software, interfaces, and enhancements needed to deliver the functional requirements.

The EMAP–IM requirements outlined in this section were defined in a series of Requirements Analysis workshops with EMAP Working Groups (see Appendix B, Data Management Needs and Practices of EMAP Working Groups). They are the basis of the system configuration described in Section 4, Technical Design.

The planning process for determining system requirements is specified in EPA Essential Elements of Information (EEI 1998) requirements. Documentation of the results of this process is presented in Appendix A, Essential Elements of Information Requirements Report.

3.2 User Needs

User needs have evolved during the course of the EMAP program. In early EMAP, the primary users were the Resource Groups at the ORD labs and other internal EPA staff. By contrast, the current program includes a diverse, distributed set of research partners and a wide audience of end users. These users range from scientists who extract sampling data sets for detailed quantitative analyses,

to members of the general public who require summary data and final reports. The basic needs of these users are similar: rapid access to relevant data. For researchers, this means the ability to electronically exchange raw data sets readily during the course of a project, or to locate data sources for integrated assessments. For end users, it may mean easy access to summarized data, such as atlases and reports.

The EMAP–IM approach and system development are designed to meet the needs of two main categories of end users: primary users and general users (see Table 3-1). Primary users are the main force behind system requirements, but they require a system that also serves most of the needs of general users.

Table 3-1. User Categories

User Category	Data Used
General Users —Non-EPA researchers and general public including resource managers (local, regional, state, federal), educational, international, insurance, legal, media	Summary and historical data
Primary Users —EMAP research partners, groups sharing EMAP research responsibilities, and EPA researchers	Preliminary, raw, summary, and historical data

3.2.1 General Users

General users represent non-EMAP researchers (federal, state, academic), government managers, policy makers, planners and resource managers, and the general public. These users have a broad set of needs, from presenting summary information for educational purposes to using final EMAP data in a newly developed model. They employ a heterogeneous suite of computing hardware, software, and networks over the Internet to access EMAP summary data and metadata that have been approved for public release.

Some examples of EMAP general users and their needs include:

- **General Public**—Users from the general public include, educational institutions, international agencies, insurance, legal, media, and others. These users have widely varying knowledge and training in information systems.
- **Planners and Resource Managers**—Planners and resource managers work at the local, state, regional, and federal level. They have a strong interest in locating summary data and reports by geographic region.

- **Researchers Outside EPA**—Scientists outside of EPA who want to use EMAP data and information in their own research need to exchange information with EMAP scientists and access completed EMAP data and documentation.

The overall needs of EMAP general users include:

- **Online interface to data**—Users need access to EMAP data and information through simple, readily available interfaces such as World Wide Web browsers. Data transfer standards such as FTP, telnet, and World Wide Web must be in place to process information requests.
- **Data integrity and accuracy**—Users need to have confidence that data made available via the EMAP–IM system are accurate and reliable and have been checked to ensure that they have been properly recorded and associated with the correct sampling events.
- **Data documentation**—Users require documentation of quality and content to determine the utility and relevance of a data set to their specific research areas. Documentation should include sampling plans, analytical and data analysis methodologies, and other assumptions.
- **Data directory**—To aid in the location of available quality-controlled data sets, EMAP–IM (AED) must maintain a central index that lists data on the EMAP–IM system or at other linked sites.
- **Data files in standard formats**—the EMAP–IM system must provide data in standard formats that a broad range of users can access (e.g., ASCII Arc/Info export files).
- **Access to EMAP summary data and reports**—Users need to access final EMAP research products that synthesize data collection and analysis efforts. These products are in standard formats that a broad range of users can access (e.g., WordPerfect, Adobe Acrobat PDF).

EMAP–IM (AED) supports these needs by providing:

- a Data Directory tool to locate data and metadata; and
- a public Web site with quality-assured data, metadata, and reports.

3.2.2 Primary Users

Primary users include the EMAP scientific community (e.g., researchers and data analysts) and organizations participating in CENR, who are responsible for management of raw data, production and transfer of summary data to EMAP–IM (AED), documentation, and data distribution. Since current EMAP studies encompass interagency partnerships, primary users include three main types:

- **EMAP Researchers**—These are the EPA and non-EPA scientists directly involved with conducting EMAP Working Group research projects (e.g., Regional staff, academic researchers, state agency and other cooperators, and other government [federal, state, local, regional] agencies). These users have scientific backgrounds but are not likely to be experts in information technology. Researchers review recently and previously collected data sets for planning and designing new field operations and analytical projects, developing ecosystem indicators or analyzing processes, and reporting. Once collected, data are assessed, analyzed, and used to produce various EMAP reports. Researchers also need to develop or use information systems that facilitate field measurement and collection of data;
- **Groups Sharing EMAP Resources and Responsibilities**—The EMAP–IM system must meet the needs of users in the CENR environmental monitoring framework, and the EPA Geographic Initiatives (e.g., Chesapeake Bay, Great Lakes, Gulf of Mexico). These users may need to use EMAP data in complementary studies. To meet these needs, EMAP has implemented data collection, management, and processing standards that are comparable to those used by these other organizations to ensure data compatibility; and
- **Other EPA Scientists and Managers**—Scientists and managers in EPA Regional offices, laboratories, and elsewhere must be able to access EMAP data for research, assessments, and management purposes. These users apply EMAP data and information in a wide variety of research and analysis activities (e.g., resource assessments). They have the same core requirements as other primary users. EPA scientists also need access to summary data and reports.

Primary users have the same core needs as those presented above for general users, with additional requirements that are listed below.

- **Data standards and guidelines**—EMAP researchers have requested standards and guidelines for management, storage, and version control for the EMAP data they collect;
- **Secure online site where EMAP researchers can share data during testing and development**—This site must support ad-hoc access to data over a secure network; and
- **Ready access to EMAP planning documents and data analysis plans, and organizing and disseminating summary results.**

3.3 Recommended Guidelines for EMAP Data Sources

EMAP depends heavily on Working Group research partners in EPA and other agencies to plan and conduct research and monitoring programs, and to collect, manage, document, and distribute data. These partners are collectively referred to as “data sources,” and are responsible for QA,

documentation, and data transfer to EMAP–IM (AED). This section includes a number of recommendations for preparation and delivery of EMAP data by researchers to ensure that high quality, documented data are available to EMAP users.

3.3.1 Types of EMAP Data Sources

EMAP data sources include EPA and non-EPA scientists who collect and create data as a result of conducting EMAP research. The major types of EMAP data sources are reviewed below; the major difference among them is the degree of control EMAP–IM has over data quality and accessibility. See Appendix B, Data Management Needs and Practices of EMAP Working Groups, for information on the data management practices of individual Working Groups.

3.3.1.1 ORD Data Sources

EMAP researchers in ORD include ORD laboratory staff that participate in EMAP Working Groups, including:

- AED—MAIA-Estuaries Working Group;
- WED—MAIA-Surface Waters Working Group, Regional EMAP/R-EMAP Working Group;
- GED—MAIA-Estuaries Working Group;
- MED—R-EMAP; and
- LV—Landscape Ecology.

3.3.1.2 Non-ORD EMAP Data Sources

EMAP data are produced by non-ORD research partners (e.g., EPA Regions, TVA, USGS, NPS, NOAA, States, and various academic institutions) who may follow their own agency’s data standards and procedures instead of EMAP’s and use a variety of data management tools (e.g., Oracle, Arc/Info, SAS).

3.3.2 Recommended Guidelines for Data Management and Delivery

The recommended guidelines for managing EMAP data are described in the following subsections.

3.3.2.1 Data Collection

Data collection is the process by which data are acquired. EMAP researchers produce data at many different levels, including new field and laboratory data, updated historical data, spatial data created through compilation and remote sensing, and aggregates of pre-existing data sets.

Requirements

Management of EMAP data in ORD Laboratories should follow existing EPA standards to the extent practical.

The majority of these researchers follow EPA standards and procedures for data collection, management, documentation, and distribution (U.S. EPA 1993). Researchers who do not follow EMAP standards should provide documentation of the methods used for QA/QC and data management. EMAP-IM (AED) works closely with ORD researchers to ensure that data are delivered according to specified standards.

Data collection methods and standards should be documented and consistent.

EMAP researchers collect data from field samples, laboratory analyses, historical data, GIS, aerial photography, and satellite images. Data are collected with a wide range of equipment and analytical methods, and are delivered in many different formats. To ensure meaningful data analysis and summaries, researchers should ensure that data collection methods follow consistent standards and that the actual processes are well documented.

QA/QC should be conducted for all data sets and metadata.

EMAP researchers should ensure that data are of known quality so they are useful for future analyses. To this end, researchers should establish and document data verification and data validation procedures. Data should be verified with appropriate data entry functions to ensure that they accurately reflect measurements, readings, observations, and analytical results. Data should be validated by comparing related data over time, and assessing data collection and processing methods. Validation is necessary to ensure that the instruments or analytical procedures are operating correctly. Inaccurate data can result in misleading research, incorrect estimation of trends, and possible misdirection of U.S. environmental policy.

Data management standards should be documented.

Researchers should document procedures and standards so that data are well understood by potential users. EMAP-IM (AED) supports researcher's efforts with guidance and assists with applying data management and delivery standards.

Data sources should ensure the quality of data made available to EMAP.

Data should be made available in a timely fashion as soon as the publication of results and data management tasks are completed.

3.3.2.2 Data Aggregates

Data aggregates include statistical data summaries derived by modifying original data through analysis, integration, or enhancement. Information management experience has shown that the effort

in aggregation management, searching, and retrieval is often equal to or greater than that for the original data collection.

Requirements

Researchers should document data aggregate methodology and sources.

In order to ensure meaningful data that are useful in the long term, researchers should document the derivation of data aggregates, including the source data sets and the methodology used to generate them.

Data aggregates should be distributed to potential users.

Researchers should make data aggregates and documentation available to other users through the EMAP–IM system. Indexing data aggregates will be a significant information challenge because large numbers of different aggregations could potentially be created.

3.3.2.3 Data Integration

Data integration is the incorporation of response, exposure, and stressor data into EMAP analyses and indices.

Requirements

Data integration methods should be documented and distributed.

Researchers should make data integration methods, results, and documentation available to other users when distributing their data.

3.3.2.4 Data Delivery—Exchange and Distribution

The basic premise of the EMAP–IM system is data sharing. For EMAP information to be valuable, it should be available to users beyond those who created or collected it. To ensure the delivery of EMAP data and information, researchers can work with EMAP–IM (AED) to provide accessible sites with documented data and information. EMAP data delivery involves two kinds of distribution mechanisms—data exchange and data distribution. **Data exchange** involves the sharing of data and information products among research partners and EMAP–IM (AED) on a rapid, ad hoc basis during testing and development. These data can be made available on a limited-access basis so that only authorized research partners can access the data. **Data distribution** involves the dissemination of quality-assured and documented data and information products to all potential users.

Requirements

EMAP data and documentation should be made available to the EMAP–IM system through appropriate repositories.

Section 3, Information Management Needs and Requirements

EMAP researchers can use the following three principal mechanisms to distribute data sets to potential users via the Internet. Few Working Groups have sufficient resources for data distribution, and most will enter them into data repositories that can relieve the burden of handling multiple data requests for commonly requested data (e.g., summary data). Options include:

- Posting data on their own publicly accessible World Wide Web server—EMAP data sets can be distributed from the web sites of agencies that collected them (e.g., USGS, NPS), or sites of individual researchers (e.g., universities, research institutes);
- Submitting data for posting on publicly accessible data repositories—Most EMAP data will be entered into established data repositories around the U.S. that specialize in different data types or geographic regions. These data repositories (such as STORET, MDN) are widely known and used by researchers in their respective disciplines. They provide stable, accessible locations for interested users to integrate and download related data;
- Transferring data to EMAP-IM (AED) for posting on the EMAP Public Web Site—EMAP-IM (AED) handles data transferred to the site.

See Tables 2-1 and 2-3 for the locations where Working Groups are planning to place their summary (final) data for distribution.

Distributed data should be accompanied by high-quality documentation and conform to Federal documentation standards.

Methodologies, indices, and other analytical tools developed by EMAP researchers are information products that should be accessible to potential users.

Methodologies, models, indices, and assessment tools developed for the program have always been one of EMAP's assets. They should be made available to potential users along with documentation. For example, various R-EMAP researchers would like to access the MAIA Habitat Quality Index and ORD indices, such as those for the environmental tolerances of specific organisms.

Researchers should follow standards that facilitate data delivery.

Data delivery standards are intended to ensure that documented data are delivered to intended recipients (other researchers, the general public) in a timely fashion and in usable formats. For example, ORD Laboratories that collect and analyze EMAP data and develop EMAP models and tools sometimes provide data in formats that the recipients cannot use (e.g., SAS, models). Data formats should be clearly indicated at download sites and in documentation. Working Groups should distribute data submission guidelines to subcontractors so they can follow the appropriate standards data.

3.3.2.5 Documentation

A major challenge in the EMAP–IM strategy is ensuring that all data and information products are accompanied by high-quality documentation. In EMAP, documentation includes metadata and Data Directory entries (data that indicates the location and general content of a data set).

Requirements

Role of documentation

EMAP researchers produce and distribute standard documentation (metadata and Data Directory entries) for all data and information products to ensure their long-term usefulness. In EMAP, documentation is most critical for new data sets collected for monitoring studies because the data may be used for many different purposes in the future.

Documentation content and format

Currently, researchers prepare documentation in a number of different formats. The EMAP documentation standard is the Data Directory and Data Catalog format, based on formats originally developed by NASA (see Section 4.4.1, EMAP Data Directory). EMAP has updated the Data Directory and Data Catalog to contain required elements of the FGDC Spatial Metadata Standard. FGDC is the leading standard and the required format for documenting spatial data produced by Federal agencies.

At a minimum, documentation should follow the EMAP standards cited above and should include data dictionaries. It would be helpful if documentation indicated the maturity of the data (e.g., raw, raw QA'd, aggregate, summary), the version of the data provided, and how the data has been updated from previous versions that users might have downloaded.

To prepare documentation that is compliant with the FGDC metadata standards, researchers can use a number of existing automated tools. Information about relevant tools is maintained by the USGS on their GCRP web site (U.S. Geological Survey 1998). EMAP–IM (AED) can assist with this task (see Section 5.6.8, Data Exchange Among EMAP Researchers).

Documentation quality

The technical level and detail included in the documentation should be adequate to answer questions about proper uses of the data.

Timing and authorship of documentation

Documentation should be completed as soon as possible after data collection and analysis to avoid loss of data and duplication of effort (e.g., through project personnel change).

Researchers who created the data should be directly involved in preparing the documentation. EMAP-IM (AED) can assist (see Section 3.4.3) with this task and should ensure that researchers have allocated adequate funds for documentation preparation. The EMAP program may need to allocate additional resources (e.g., funding, staff) to ensure completion.

EMAP Data Directory entries as documentation

Taken together, the EMAP Data Directory and Data Catalog provide a very complete format for documenting EMAP data that has incorporated key elements (e.g., locational accuracy, related citations, distributions, and graphic files) of the FGDC requirements. Data Directory entries are considered essential documentation of data sets that allows users to locate them and their accompanying documentation. Entering Data Directory information (e.g., data types, keywords, contact names) about a data set should be a minimum documentation requirement for all data sets produced by EMAP researchers. Data sources should follow the EMAP standard format outlined in Frithsen and Strebel (1995b) and updated in an EPA addendum (U.S. EPA 1996g). EMAP-IM (AED) can assist EMAP researchers with this task.

3.3.2.6 Data Archiving

Data archiving is the activity of making data backups at all stages of data maturation (raw, processed, aggregate, summary) for long-term storage.

Requirements

EMAP researchers should establish archiving procedures for all data.

Researchers should establish archiving procedures to ensure no loss of data. EMAP-IM (AED) can provide assistance with this data management task. Standard approaches are being developed by SIMCorB which are consistent with ORD requirements.

3.3.2.7 Data Storage

Data storage is the long-term maintenance of data.

Requirements

Researchers should ensure long-term data storage.

The majority of EMAP data will be maintained for the long term by the researchers that created it. Subsets of these data—summary data sets, data products, and selected original data sets—will be made available through data repositories. For more information on where Working Group data will be stored, see Tables 2-1 and 2-3 and Appendix B, Data Management Needs and Practices of EMAP Working Groups.

3.3.2.8 Version Control

Data will change over time as errors are corrected and new observations are added, as multiple data sets are integrated into data aggregates, or when cooperating researchers generate multiple versions of a data set. These versions should be tracked so that users can understand the data they are accessing.

Requirements

Researchers need to track versions and relationships among data sets that they collect, modify, integrate, and distribute. They may also need to develop mechanisms for notifying users of data set updates and version changes. Version control methodology is not currently available in most EMAP Working Groups. EMAP can provide guidance with this activity.

3.4 EMAP–IM Functional Requirements

Functional requirements include those capabilities and components that EMAP–IM (AED) must provide to ensure the flow of data from data collectors to users. Most of the responsibility for functional requirements resides with EMAP–IM (AED), but Working Groups and the IMWG share a number of tasks, particularly in relation to data distribution. The following subsections describe the components, capabilities, and guidance that are needed to fulfill these responsibilities.

3.4.1 Track EMAP and Non-EMAP Data Relevant to EMAP Research

The EMAP–IM system must track a wide range of EMAP and non-EMAP (i.e., external) data sets including summary data sets, data aggregates, historical data sets (e.g., historical monitoring data collected by sewer authorities), methodology (e.g., Benthic Index, statistical models), information products (e.g., MAIA landscape atlas, reports), documentation, and GIS data (LANDSAT satellite images, Landscape Ecology indicator coverages, EPA River Reach RF3 data).

3.4.1.1 Provide Access to Individual EMAP Data Products by Tracking Them at the Widely Dispersed Data Distribution Sites

The EMAP–IM system must function as a central directory for locating EMAP data products and documentation at data distribution sites widely dispersed across the Internet. The primary challenges are: to convey the complexity of relationships between data and its corresponding metadata; to know the derivation of data aggregates and their associated data sources; and to track numerous data types from remote sensing images to publications to clearinghouses to relational databases. In addition, a variety of different types of access mechanisms will be used to deliver the data (e.g., print, online data clearinghouses), and the EMAP–IM system must be compatible with these and adapt to them when they change.

The Data Directory and Public Web Site must track and cross-reference individual data products down to the level of individual data sets so that relationships among related data sets can be understood (e.g., data sets collected in the same study but stored in different data repositories; relationships between data and its associated documentation; the derivation of data aggregates from their associated data sources). For example, a single R-EMAP project can contain many small data sets from a number of different disciplines (e.g., atmospheric deposition of mercury, fish community integrity in streams, remote sensing landscape indicators). These data sets will be archived in multiple independent repositories (air data in the MDN or EPA AIRS database, water data in the STORET database). EMAP-IM system users will need to be able to bring together these related data.

3.4.1.2 Provide Documentation of the Quality of External Data Sources and Metadata

The EMAP-IM system needs to provide its users with some indication of the status and quality of external data sets. EMAP researchers have requested assistance with locating, acquiring, and using some external data, since their quality, accessibility and suitability for particular EMAP analyses can vary widely. The documentation may be insufficient to convey these aspects of the data. For example, many groups want to use the EPA RF3 data as a baseline for sampling, but find that the quality varies widely by region and the data must be modified to make them usable. EMAP-IM can track the status and locations of data sets that are frequently used by EMAP researchers (e.g., EPA RF3, USGS hydrography, etc.) by citing the location of their documentation in the Data Directory. The EMAP Data Directory and Public Web Site can point to data sets useful to EMAP researchers and provide hypertext links to web sites where the data and its documentation can be obtained.

3.4.1.3 Organize Data and Metadata for Ease of Retrieval and Updating

Tracking EMAP data includes locating and organizing references to data and metadata, much of which is not currently indexed in any directories. The Data Directory must also handle listings for a wide variety of information types, including field monitoring data sets, documentation, other Data Directory entries, raw data collected by EMAP laboratories, and summary data from EMAP collaborators in other organizations. Directory entries must include contact names or hyperlinks (to the EMAP Public Web Site for EMAP data; or external web sites for non-EMAP data). The system must also provide access to data in many different formats (e.g., data clearinghouses, academic web sites, Internet FTP, libraries), and be compatible with the many types of technologies used at those sites. This access can include other online agency data directories, particularly those associated with CENR, particularly the GCRP web site (GCRP 1998).

3.4.1.4 Keep Data Directory Entries Up-to-Date and Synchronized with Evolving Versions of EMAP Data and Data Sources Frequently Accessed by EMAP Users

Tracking multiple versions of EMAP data sets and documentation in the distributed EMAP-IM system will be a critical issue in EMAP data management and documentation. Some form of version control methodology may be needed to track entries in the Data Directory, Data Catalog, and data repositories. Version control issues include: keeping distributed files synchronized, allowing for easy

update while maintaining integrity; keeping track of data changes; and ensuring that metadata is updated when data sets change. Data sets must be tracked when they are updated or modified so that multiple versions can be distinguished and their relationships understood. In addition, they must ensure that the accompanying documentation is current and accurate for each data set. For non-EMAP data sets that EMAP users access frequently, EMAP-IM could explore methods for tracking updates and revisions.

3.4.2 Facilitate Rapid, Ad Hoc Data Exchange Among EMAP Researchers

Since EMAP research responsibilities are shared by EPA and non-EPA research partners at many different sites, EMAP-IM must provide efficient mechanisms for sharing data sets and information products among sites for information that is not yet published on the Public Web Site. The chief purpose of this sharing is to allow testing and development of information products by EMAP-IM (AED) and the researchers so they can determine if the material has passed criteria for being placed on the Public Web Site (U.S. EPA 1998a). The steps EMAP-IM (AED) will follow to address these requirements are discussed in Section 5.6.8, Data Exchange Among EMAP Researchers.

3.4.3 Provide Standards, Tools, and Support to Users and Data Collectors

The following subsections summarize the role of EMAP-IM in providing guidance and assistance to researchers with tools and standards, research planning, data management, documentation, and delivery.

3.4.3.1 Maintain and Disseminate Standards for Data Collection, Management, Documentation, and Distribution

EMAP-IM (AED) must make existing standards available to EMAP data collectors for:

- data collection;
- codes for QA, taxonomy, chemistry, and other parameters;
- data management;
- data submission;
- documentation; and
- version control methodology with procedures for notifying users of version changes.

One important area in which EMAP-IM must maintain and distribute standards is for metadata. Researchers employ a variety of methods for preparing documentation to meet the EMAP metadata standard. EMAP-IM (AED) maintains the EMAP standards and can assist with documentation tasks.

3.4.3.2 Support EMAP Data Collectors and Users with Data and Metadata Preparation

EMAP-IM (AED) can provide direct assistance to EMAP researchers using the tools and standards outlined above. In addition, EMAP-IM (AED) can provide assistance with completing documentation and Data Directory entries.

EMAP-IM (AED) can also provide researchers with data management guidance if they do not have standard data management methods and software (e.g., Oracle). For example, several groups have indicated that they need this assistance because they currently manage data in analysis packages like Arc/Info and SAS which have limited data management capabilities.

3.4.3.3 Participate in CENR Standards Development and Implementation

EMAP's participation in the CENR monitoring framework requires compliance with a number of data and metadata standards that CENR adopts for data documentation, transfer, formats, and policy. The main purpose of these standards is to ensure maximum interoperability with the data from other Federal agencies. The standards most closely follow the GCDIS. The GCDIS "is the set of individual agency data and information systems supplemented by a minimal amount of cross-cutting new infrastructure, and made interoperable by use of standards, common approaches, technology sharing, and data policy coordination." (CENR 1994). For further discussion of standards development, see Section 6.7.2, Tasks.

3.4.3.4 Distribute EMAP Tools and Automated Procedures for Research Planning and Implementation, and Monitoring Network Design

EMAP-IM must continue to make existing tools available for planning and implementing monitoring programs. A number of EMAP Working Groups (e.g., internal ORD researchers) use the early program tools, including:

- procedures for determining the station locations to be sampled and the types of samples collected at each location;
- automated data collection methodologies;
- methods for managing data from the stations; and
- site-data handling protocols for field data collection.

Use of these tools is at the discretion of the researchers, who can ask EMAP-IM (AED) for assistance. Some Working Groups have requested that such assistance include guidance to these tools for ease of understanding, such as in "Lessons Learned" synopses of successful approaches and SOPs.

3.4.4 Maintain and Update EMAP–IM System (Components and Network Connections)

EMAP–IM must maintain the system resources, including the Data Directory, Public Web Site and Internal Web Site, Data Catalog, links to distributed sites, and other information resources that are more fully described in Section 3.5, System Requirements, and Section 4, Technical Design.

EMAP–IM, through the ORD Laboratory Divisions and with support from the Office of Information Resource Management’s (OIRM) Enterprise Technology Services Division (ETSD), will stay abreast of new technological developments in the hardware and software used (Oracle, SAS, Arc/Info, World Wide Web) and in advances made by others in managing scientific databases. These new capabilities will be applied to the EMAP system when practical.

3.4.5 Deliver EMAP Data and Information Managed by EMAP–IM (AED)

A few data sets are managed and distributed by EMAP–IM (AED), including the Estuaries Resource Group and MAIA-Estuaries data, data from EMAP researchers that request data storage (e.g., R-EMAP); and relevant non-EMAP orphan data sets that have no other repository. EMAP–IM (AED) follows EMAP and applicable Federal standards in order to ensure long-term data availability, quality, and integrity.

3.4.5.1 Improve Access to EMAP Resource Groups Data Sets and Metadata

Data sets from 1990–1995 are being maintained by the Resource Groups, but summary data and documentation files are being made available to users through the EMAP Public Web Site. Several data sets are still unavailable and are needed by the Working Groups, so it is important for EMAP–IM (AED) to continue working to deliver these data to users.

3.4.5.2 Capture “Orphan” Data Sets in the EMAP–IM System

The EMAP–IM system primarily tracks data distributed at other sites, but in cases where orphan data sets (data sets that are not being actively maintained or have no other long-term repository) are useful to EMAP, they could be managed, documented, and distributed through the EMAP–IM system. These data sets can include those from EMAP researchers as well as those from external researchers. This activity would require allocation of EMAP resources for managing and distributing the data.

3.5 System Requirements

System requirements are the software, hardware, and network configuration that must be in place to provide the functionality specified in this section, including:

- tracking distributed data sets;

- improving the flow and delivery of EMAP and related data;
- increasing the accessibility of the EMAP Data Directory to queries through the Internet;
- increasing interoperability with other information systems (e.g., Consortium for International Earth Science Information network [CIESIN], EPA's ReVA); and
- allowing data and information exchange among the heterogeneous set of EMAP researchers using diverse hardware, software, and networks.

3.5.1 Overall System Design

The EMAP-IM system must be based on open standards that are flexible and responsive enough to serve existing program needs but also adapt to:

- a constantly expanding and changing user base;
- an expanding set of the types and quantity of available data; and
- new hardware, software, and infrastructure that can enhance system effectiveness.

The challenge is to update the system in response to these changing needs.

3.5.2 System Components

The EMAP-IM system components must organize available data and metadata into structures and systems that promote the efficient maintenance and location of pertinent information. This organization is accomplished through a combination of directory structures, relational databases, and web site organization. Requirements for meeting these objectives are outlined below.

3.5.2.1 EMAP Data Directory

The Data Directory must provide information about general content, location of data, and contact information.

Requirements

Functionality

The EMAP Data Directory must provide a central index that helps users locate collected EMAP data and links to where those data can be accessed.

Standards

Standards for the Directory are based on early EMAP standards (Frithsen and Strebel 1995, and U.S. EPA 1996f), and must follow FGDC and Global Change Master Directory (GCMD) requirements.

This coordination will enable the Data Directory to be part of a network of environmental data providers (e.g., EIMS, CIESIN). Standards that may be adopted include the Z39.50 protocol.

Format

The Data Directory must be maintained in the Oracle relational database management system (Oracle RDBMS) on the EMAP Internal Web Site.

Accessibility

The EMAP Data Directory must be available through the EMAP Public Web Site.

3.5.2.2 EMAP Data Catalog

The Data Catalog must provide metadata about data sets on the EMAP Web site.

Requirements

Functionality

The EMAP Data Catalog must provide useful information so that the data can be correctly interpreted and used.

Standards

The Data Catalog must conform to EMAP requirements (Strebel and Frithsen 1995b; U.S. EPA 1996h), which have recently been made compatible with FGDC metadata requirements.

Format

The Data Catalog should be updated to use standard metadata formats to provide interoperability with other federal agency data catalog standards, particularly those associated with CENR.

Accessibility

Existing Data Catalog files for EMAP data must be available via the EMAP Public Web Site.

3.5.2.3 EMAP Public Web Site

EMAP's Public Web Site must provide a publicly accessible interface for EMAP data and information (Data Directory, Data Catalog, bibliography, online publications, and links to data sets).

Requirements

Functionality

The EMAP Public Web Site must make EMAP data and information available to all World Wide Web users.

Standards

The EMAP Public Web Site must be maintained according to the standards for delivery of EMAP data and information outlined in Strebel and Frithsen (1995a) and the EPA addendum (U.S. EPA 1998b). EMAP policy states that no data files can be placed on this site unless accompanied by metadata (e.g., Data Catalog) files.

Format

The Public Web Site must be in formats that can be read by World Wide Web browsers (e.g., HTML, databases compatible with web servers). The components on the site are formatted as follows: the Data Directory is presented in HTML format, the Data Catalog files are in ASCII text format, and available data files are in ASCII and Arc/Info export formats.

Accessibility

The Public Web Site must be freely accessible to all users through Web browsers.

3.5.2.4 EMAP Internal Web Site

The EMAP Internal Web Site must provide a location for sharing information within the program for development and testing before it is placed on the EMAP Public Web Site.

Requirements

Functionality

The EMAP Internal Web Site must serve the EMAP community's need for:

- development and testing of EMAP data and tools; and
- rapid, ad hoc exchange of EMAP data and information.

Standards

The EMAP Internal Web Site must follow internal EPA/ORD standards and provide adequate security for data under development.

Format

The Internal Web Site components should be in formats consistent with EMAP needs and standards (e.g., Data Directory in Oracle, Data Catalog files in ASCII text, data files in ASCII and Arc/Info export formats).

Accessibility

The Internal Web Site must be made accessible to all researchers preparing EMAP data for sharing within the program and distribution to end users.

3.5.3 System Configuration—Software, Hardware, Network, and Online Resources

The configuration of the EMAP–IM system consists of connections between AED servers and the Internet, the public access server at RTP, and the software and data files on the AED servers (see Section 4.5, System Configuration). Software and hardware used in EMAP must fit into the EPA computing environment but be flexible enough to incorporate new interfaces and formats.

The EMAP–IM system architecture must be kept sufficiently flexible to allow use of current technology yet be capable of adapting to future technology.

3.5.3.1 Software and Data Processing Resources

Data processing resources include the software used to collect, analyze, store, and distribute data. These resources cover a wide range of packages that are distributed among EMAP data sources and EMAP–IM.

Requirements

Database Management Tools

EMAP–IM (AED) and the researchers currently use a combination of tools to manage data and metadata. The Data Directory is managed in the Oracle RDBMS on the Internal Web Site, and access is provided through the Oracle web server option in both the internal and public web sites. Researchers maintain EMAP data sets using a number of tools, (e.g., SAS and Oracle), and they convert the files to ASCII for placement on the Public Web Site. Data Catalog files are managed internally as WordPerfect, but distributed as ASCII or HTML files.

User Interfaces

EMAP–IM system user interfaces must simplify data access, offer the best access based on existing technology, and adapt to new interface technology which is rapidly evolving. Working Groups indicated that it is important that the EMAP–IM system allow the user to access data through simple interfaces without having to understand the underlying complexity of the data.

3.5.3.2 Hardware Resources

The EMAP–IM system uses standard EPA servers at AED to maintain the Data Directory, Data Catalog, and the EMAP Internal Web Site. They transfer Data Directory, Data Catalog, data sets, and other information ready for public release to the Public Web Site.

EMAP researchers need access to on-site, networked workstations with a wide range of capabilities.

Requirements

Capacity

Working Groups and EMAP–IM (AED) must have access to sufficient computing capacity to work with EMAP data types and exchange information (Internet, data files) with other EMAP researchers. See Section 4.5 for a summary of the existing system capacity at AED; see Appendix C, Inventory of EMAP Data, for estimates of data volumes to be stored by Working Groups.

3.5.3.3 Network Availability and Capacity

The EMAP–IM system uses the standard EPA network configuration for its Internet connections and formats.

Requirements

Online Resources

The EMAP–IM system must provide researchers access to data under development. This site must be:

- able to provide access to all research partners;
- available on an ad hoc basis for data upload and download; and
- able to handle files as large as 1 gigabyte (GB).

The existing Internal Web Site can be expanded to provide this capability (see Section 5.6.8, Data Exchange among EMAP Researchers).

Links to Outside Resources

To provide the access to data required by EMAP researchers, the EMAP–IM system must provide the capacity to exchange information with a wide variety of computer types, software packages, and file systems without problems due to incompatibilities or versions.

3.5.3.4 Storage and Distribution

EMAP–IM requires reliable storage and retrieval systems for large amounts of data (see Appendix C, Inventory of EMAP Data, for estimates of data volumes to be stored by Working Groups).

Requirements

Reliable Storage Media

EMAP–IM stores data on the AED servers and on appropriate storage media, including tapes and CD-ROMs.

EMAP researchers will provide reliable storage for their data in a variety of ways. Some groups have access to ORD servers that can store large amounts of data (e.g., DISPro UVB data on the NERL pages); others have limited resources and rely on data repositories such as EMAP-IM or STORET.

3.5.3.5 Data Security

Data security involves maintaining the integrity of the data and ensuring that it is properly stored so it cannot be corrupted. Security is the responsibility of those managing and distributing the data (from EMAP to data clearinghouses to individual researchers).

Requirements

Protecting data from damage

Maintaining data and protecting it from damage is the responsibility of the data source (data owner). EMAP-IM is responsible for data maintained on its servers and research partners are responsible for the data.

Limiting distribution of certain data

Some data sets are not accessible in complete form at all times to all users because of confidentiality (e.g., sampling locations, property owner names) or because data are still being analyzed and used to publish results. Working Groups and EMAP-IM (AED) ensure appropriate distribution of confidential data. Data could be distributed in limited versions (e.g., forest stand monitoring results could be released without exact locations of sample site locations).

Delayed release of data

Data collectors may have concerns about the timing of data release based on their need to publish. In these cases, a hold-back period before public release can be used when requested.

3.6 Conclusions

The user needs, data source requirements, EMAP-IM functional requirements, and system requirements are partially fulfilled in the existing system (see Section 4, Technical Design). Future enhancements can address remaining gaps (see Section 6, Implementation Plan). To ensure the success of the program, EMAP must have a long-term commitment to maintaining and upgrading the system to ensure that data, metadata, and information are retrievable and meaningful for future users.

Section 4 Technical Design

- 4.1 Purpose**
- 4.2 Background of EMAP–IM System Development**
 - 4.2.1 Early EMAP Information Management System (1990–1995)**
 - 4.2.2 Current EMAP Information Management System (1996–)**
- 4.3 System Concept And Overview of Technical Structure**
- 4.4 System Components**
 - 4.4.1 EMAP Data Directory**
 - 4.4.1.1 Purpose**
 - 4.4.1.2 Background**
 - 4.4.1.3 Design**
 - 4.4.1.4 Entries**
 - 4.4.1.5 Standards**
 - 4.4.2 EMAP Data Catalog**
 - 4.4.2.1 Purpose**
 - 4.4.2.2 Background**
 - 4.4.2.3 Design**
 - 4.4.2.4 Entries**
 - 4.4.2.5 Standards**
 - 4.4.3 EMAP Public Web Site**
 - 4.4.3.1 Purpose**
 - 4.4.3.2 Background**
 - 4.4.3.3 Design**
 - 4.4.3.4 Standards**
 - 4.4.4 EMAP Internal Web Site**
 - 4.4.4.1 Purpose**
 - 4.4.4.2 Background**
 - 4.4.4.3 Design**
 - 4.4.4.4 Standards**
 - 4.4.5 EMAP Summary Data Sets**

- 4.4.5.1 Purpose
- 4.4.5.2 Background
- 4.4.5.3 Design
- 4.4.5.4 Standards
- 4.5 System Configuration
- 4.6 EMAP Archival Plan
- 4.7 System Evaluation
 - 4.7.1 Data Accessibility
 - 4.7.2 Flexibility of Design to Adapt to Future Technological and Program Changes
 - 4.7.3 User Satisfaction
 - 4.7.4 Benefits and Costs
 - 4.7.5 Risks and Contingencies
- 4.8 Need For System Enhancement
- 4.9 Conclusions

The EMAP–IM system provides components that facilitate timely access to relevant data and information for the diverse set of EMAP users. The system has been adapted to the needs of the current EMAP program by enhancements to the Data Directory and EMAP Public and Internal web sites that allow EMAP–IM to manage EMAP scientific information and distribute it through simple interfaces.

4.1 Purpose

Because EMAP data management is distributed at many locations across the Internet, most information management requirements (see Section 3) cannot be automated in a centralized system. Instead, EMAP–IM system components provides a way to link the distributed capabilities of the EMAP community’s database applications, hardware, and data (Figure 4-1). This section describes the combination of features (system components, software, hardware, and network connections) that have been implemented to meet these requirements and support a national program over an extended period. The system is designed to:

- meet the needs of a constantly expanding and changing user base;
- accommodate significant changes in the types, quantity, and location of data being used and collected; and
- incorporate new technology (hardware, software, infrastructure) that can enhance the system’s ability to serve users and interact with other data systems.

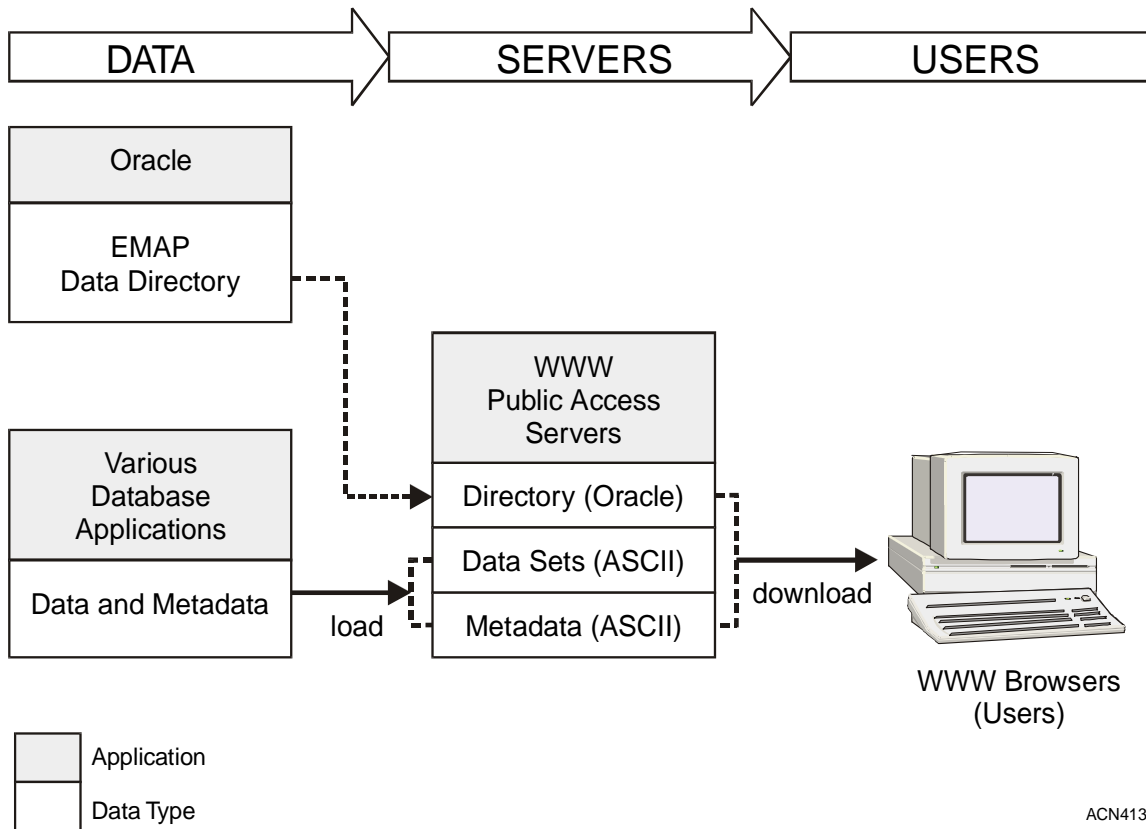


Figure 4-1. Components of the distributed EMAP-IM system.

4.2 Background of EMAP-IM System Development

4.2.1 Early EMAP Information Management System (1990-1995)

The EMAP Resource Groups began collecting data in 1990, before a central information management group was formed. As a result, each Resource Group developed independent data management systems and policies. In general, the data were entered into SAS, as well as in commercial and custom field computer packages; a few were maintained in relational databases (Oracle). Most Resource Groups were based at ORD labs and followed standard EMAP sampling design and data collection methodologies; some data sets were collected by groups outside of EPA (Forests, Agroecosystems). Resource Groups distributed data to users by request. Several Resource Group data sets (e.g., Estuaries, Surface Waters, Great Lakes) have since been made available on the EMAP Public Web Site and Internal Web Site (on the EPA Intranet/internal wide area network), but a number are still available only by request to the Resource Group that collected them.

In the early program, Central EMAP–IM was the coordinating group that oversaw all information management in EMAP. Its challenges were to integrate the results of the independent Resource Group data systems, and conduct information management research and tools development. In 1993, EPA obtained Oracle as the agency relational DBMS. Central EMAP–IM developed an information management plan and relational database management system based on a client-server model in which each Resource Group’s data would be a node linked to the Central EMAP–IM node. During 1993–1994, Central EMAP–IM used Oracle CASE tools to design an Oracle database that could store data from many different scientific disciplines. The design included a Data Directory, Data Catalog, and centralized database for EMAP data (U.S. EPA 1994a).

The function of the EMAP Oracle database (U.S. EPA 1994b) was to hold the Data Directory and Resource Group data. The Data Directory pointed interested users to relational databases; included details about the size, accessibility, and the pedigree of available data; and included links to the Data Catalog. The Data Catalog provided documentation files (in text format) that contained detailed information about each data set. The database portion of the EMAP–IM system consisted of Oracle relational tables that stored EMAP data sets, accompanied by a Data Dictionary that listed the individual fields or columns (attributes) in each table (entity). The system was documented with Oracle CASE tools; documentation includes data dictionaries (for the Data Directory, see U.S. EPA 1996g) and an entity relationship diagram. Central EMAP–IM also developed standards for Resource Group data collection, management, documentation, and other activities; some of these standards are being actively used in the current system (see Appendix F, Overview of EMAP Information Management Policies, Guidelines, and Standards).

4.2.2 Current EMAP Information Management System (1996–)

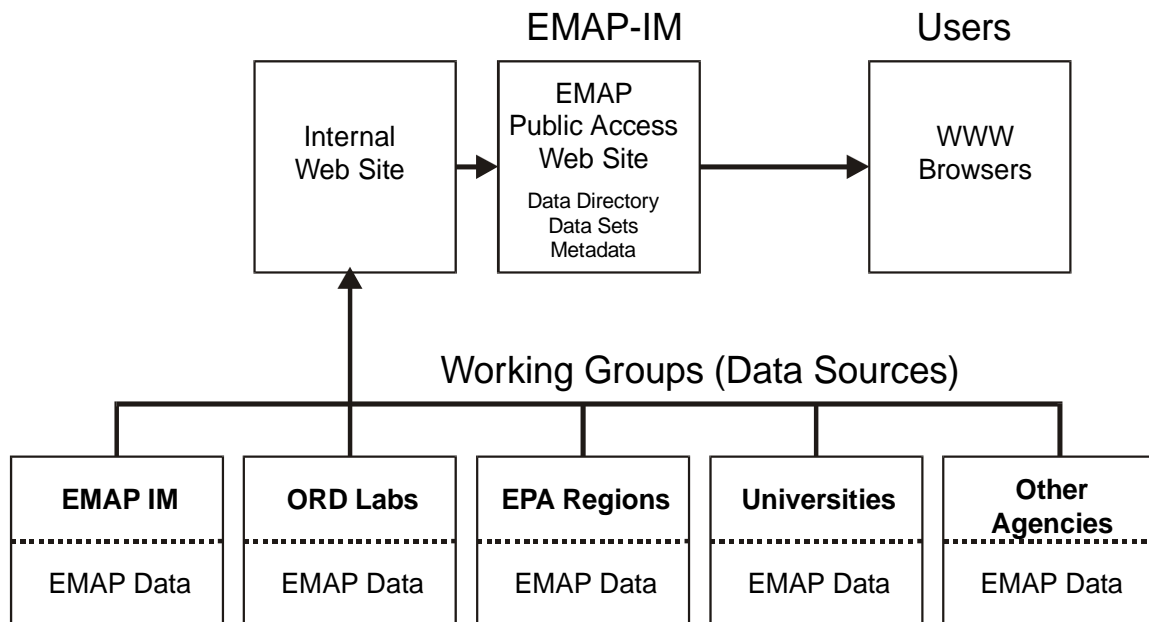
The primary focus of current EMAP–IM system development is to upgrade the components to enhance their functionality; improve their accessibility on the Internet, and create simple interfaces to the complex network of data relevant to EMAP. This section describes the technical design of the system.

4.3 System Concept and Overview of Technical Structure

The current EMAP–IM system is an Internet-based directory founded on a central Data Directory and Public Web Site. Together, these tools provide access to reliable, cross-referenced information about the location and utility of data and metadata relevant to EMAP. This system currently serves the needs of a broad, diverse set of users. It is a continuation of the distributed data management approach used in the early (1990–1995) EMAP program (Shepanek 1994, U.S. EPA 1996a), in which summary data flows from decentralized research groups to a central coordinating site (Figure 4-2). This approach is similar to that used by many other research and monitoring organizations, such as the National Oceanic and Atmospheric Administration (NOAA) and the NSF Long-Term

Ecological Research Program (LTER 1995). Many of the standards used to create and maintain the components have evolved from the U.S. Global Change Research Program (GCRP 1995a, 1995b).

The technical approach outlined here for EMAP-IM system development and implementation includes both infrastructure (hardware, software, networks) and personnel that maintain the infrastructure. The system is based on existing EMAP standards (Frithsen and Strebel 1995, Strebel and Frithsen 1995a, Strebel and Frithsen 1995b, NASA, 1991) for data documentation and distribution. The overall system structure is modeled after ORD Division laboratories data management structures, which include a Data Directory, data sets, and metadata files.



ACN414

Figure 4-2. Flow of data and metadata from data sources to EMAP Word Wide Web Site.

4.4 System Components

The EMAP-IM system includes the following components upgraded from the early EMAP program:

- Data Directory;
- Metadata (Data Catalog) files;
- Summary data sets (from EMAP Resource Groups and Working Groups);
- Public Web Site (RTP); and
- Internal Web Site (AED).

Section 4, Technical Design

These components are managed by EMAP-IM (AED) and content is provided by researchers in the Resource Groups and Working Groups. The types of EMAP data and the information management system component used to manage these data are shown in Table 4-1.

Table 4-1. Types of EMAP Data and the Corresponding EMAP System Components Used to Manage Them

DATA TYPE	DATA DIRECTORY	DATA CATALOG	EMAP DATABASE & WEB SITES	ORD DIVISIONAL DATABASES	NON-EMAP DATA REPOSITORIES
Early EMAP (1990–1995) data	X	X	X	X	X
Current EMAP (1996– present) data					
EMAP–collected	X	X	X	X	X
EMAP–funded grants program	X				X
R-EMAP–projects	X	X	X	X	X
External (non-EMAP) data and metadata	X				X
Broadly useful to or modified by EMAP	X	X	X		

4.4.1 EMAP Data Directory

4.4.1.1 Purpose

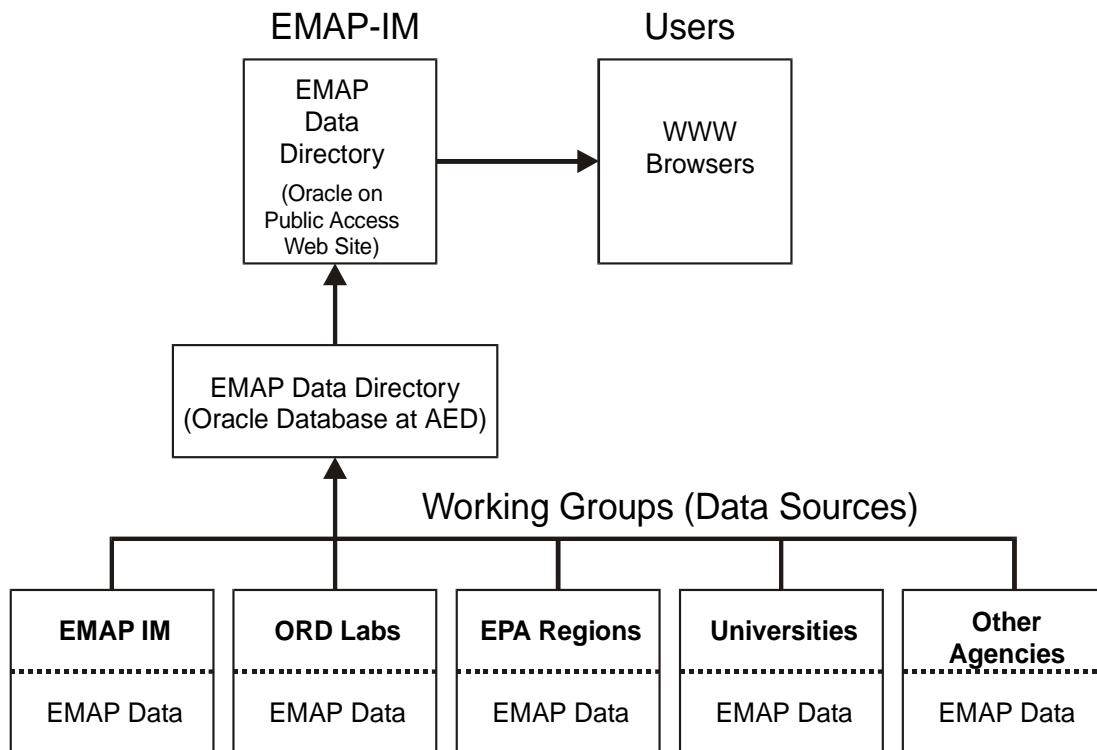
The EMAP Data Directory allows users to locate data of interest by providing information about the location and accessibility of data sets (e.g., geographic information system (GIS) coverages, spreadsheet tables, database files, remote sensing images). EMAP researchers prepare Data Directory entries for the data they collect. (Researchers can suggest to EMAP-IM non-EMAP data sets they would like cited in the Data Directory (e.g., U.S. Geological Survey hydrography data), but in most cases these external data will not be acquired, just listed.) The flow of Directory entries from researchers to EMAP-IM (AED) is shown in Figure 4-3.

4.4.1.2 Background

The original EMAP Data Directory was implemented in SAS, and Resource Groups have been entering information since 1991. In 1994, it was converted into tables in the Central EMAP-IM Oracle database. The format was based on the National Aeronautics and Space Administration (NASA) Directory Interchange Format (DIF) (NASA 1991).

EMAP Resource Groups were responsible for creating Data Directory entries for data they collected. EMAP developed a prototype Oracle data entry tool for this purpose, but only a few Data Directory entries were created.

In 1995, EMAP-IM (AED) extracted the Data Directory tables from the Central EMAP-IM Oracle database. The Oracle Forms query tool—originally developed to assist with maintenance of the overall database—was revised to be specific to Data Directory management, and was upgraded to Oracle Forms 4.5. The Directory now contains a partial listing of early EMAP data, and the remaining entries are being added by the Resource Groups to create a comprehensive inventory.



ACN415

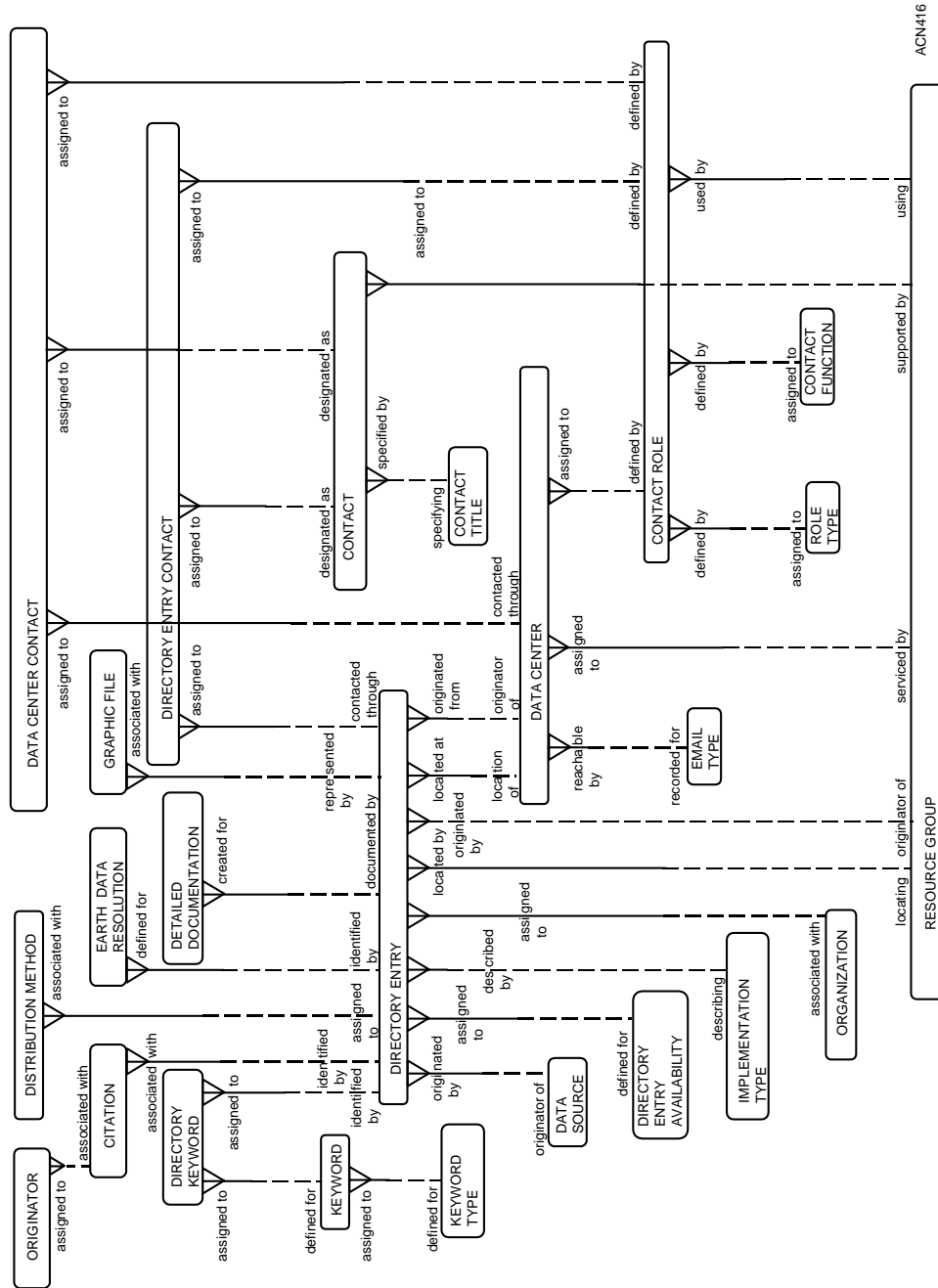
Figure 4-3. Flow of Data Directory information from data sources to EMAP World Wide Web.

EMAP-IM has developed an automated tool for creating Data Directory entries. This tool is based on the NASA DIF Writer and includes built-in checks for required fields. It is currently available on the EMAP Internal Web Site.

4.4.1.3 Design

The EMAP Data Directory is maintained by EMAP-IM (AED) in the format and standards based on the Data Directory tables extracted from the early EMAP Oracle database (Frithsen and Strebel 1995, U.S. EPA 1998b). The Entity-Relationship Diagram for the Data Directory is shown in Figure 4-4; the Data Dictionary is presented in Table 4-2. The data dictionary for the Directory can be found in U.S. EPA (1996g). The Directory tables hold information about data collected by or relevant to EMAP research. Data sets are included as they become available on the EMAP Internal and Public Web Sites.

Section 4, Technical Design



ACN416

Figure 4-4. Entity Relationship Diagram for EMAP Data Directory.

Section 4, Technical Design

Table 4-2. List of All Attributes in the Directory Database

Data Set Identification		
• Organization Name:	• Version Number:	• Data Set Progress:
• Sub-organization Name:	• Entry Date:	• Data Set URL:
• Data Set ID:	• Revision Date:	
Data Set Description		
• Data Set Name:	• Data Set Creation Date:	• General Keyword: (Multiple entries allowed)
• Data Set Format:	• Data Set Revision Date:	• Data Set Comments:
• Data Set Source:	• Abstract:	
• Number of Sampling Stations:		
Data Quality Comments		
• Data Quality Comments:		
Temporal Period		
• Sampling Start Date:	• Sampling Start Date - Month:	• Sampling End Date - Month:
• Sampling End Date:	• Sampling End Date - Year:	• Sampling Frequency:
• Sampling Start Date - Year:		
Geographic Coverage		
• Minimum Latitude:	• Minimum Altitude:	• Maximum Depth:
• Maximum Latitude:	• Maximum Altitude:	• Depth Units:
• Minimum Longitude:	• Altitude Units:	• Locational Keywords:
• Maximum Longitude:	• Minimum Depth:	• (Multiple entries allowed)
Data Center		
• Data Center Name:	• State:	• EMAIL Address:
• Address1:	• Zip Code:	• Preferred Contact Position:
• Address2:	• Country:	• Originating Organization:
• Address3:	• Voice Phone:	• Originating Sub-organization:
• Address4:	• FAX Phone:	• Originating Data Center:
• City:		

Section 4, Technical Design

Table 4-2. Continued

Contacts (Multiple entries allowed)		
• Contact Title:	• Address2:	• Zip Code:
• Contact Last Name:	• Address3:	• Country:
• Contact First Name:	• Address4:	• Voice Phone:
• Contact Middle Initial:	• City:	• FAX Phone:
• Contact Role:	• State:	• EMAIL Address:
• Address1:		
Data Set Citation (Multiple entries allowed)		
• Originator:	• Publication Date:	• Edition:
• Title:	• Publication Place:	• Data Presentation Form:
• Series Name:	• Publisher:	• Citation URL:
• Issue Identification:		
Earth Data Resolution		
• Latitude Resolution:	• Altitude Resolution:	• Depth Resolution:
• Longitude Resolution:	• Altitude Units:	• Depth Units:
Browse		
• File:	• Caption:	• Graphic File URL:
• Description:	• Format:	
Distribution		
• (Multiple entries allowed)	• Distribution size:	• Fees:
• Distribution Media:	• Distribution Format:	

4.4.1.4 Entries

Data Directory entries are being created for EMAP data sets through a combination of effort by researchers and EMAP-IM. Entries contain information about the following data:

- early and current EMAP data and metadata;
- external data sets that EMAP researchers identify as broadly useful; and
- external status tables on web site to track data sets.

The Data Directory has been expanded to track many different data types, including spatial data. Spatial data useful to EMAP includes a number of sources frequently used by EMAP researchers, including:

- Multi-Resolution Land Characteristics data (U.S. EPA 1998h), and other remote sensing data at the Earth Resources Observation Systems (EROS) Data Center;
- EPA sites that track the availability of GIS data, including Surf Your Watershed (U.S. EPA 1998i); and
- the EPA Geographic Information Systems Tools web site (GISTools 1998).

4.4.1.5 Standards

The Data Directory entries follow the standards and formats developed in the early EMAP program and updated for the current program (Frithsen and Strebel 1995, U.S. EPA 1996f). The format of the Directory and Catalog are compatible with the Global Change Master Directory (GCMD) operated by NASA (NOAA 1996). The EMAP-IM system contains fields in the Data Directory and Data Catalog that allow these two components to meet FGDC standards (FGDC 1994, NOAA 1996, U.S. EPA 1996g). The compatibility of these components with GCMD and FGDC contributes to the EMAP goal of sharing information with other agencies under the CENR monitoring framework.

4.4.2 *EMAP Data Catalog*

4.4.2.1 Purpose

The EMAP Data Catalog provides users with detailed documentation (metadata) so that they can understand, correctly interpret, and use data files. It provides much more detail about the origin and quality of a data set than the Data Directory, which is primarily designed to help users locate data. The Data Catalog is maintained separately from the Data Directory and only contains entries for data managed by EMAP.

4.4.2.2 Background

The original EMAP Data Catalog was designed in 1994 and was intended to be implemented in Oracle¹. However, EMAP instead stored the information in text files (WordPerfect) based on NASA's GCMD format (Strebel and Frithsen 1995b). A few Data Catalog files were also loaded to Oracle Book, a hypertext-based product. Data Catalog files have been created for most of the early EMAP data sets.

4.4.2.3 Design

The Data Catalog includes information about a data set that includes information about the scientific and data management manipulation of the data, quality control/quality assurance, data accessibility, and other details. A Data Dictionary for the Data Catalog can be found in Strebel and Frithsen

¹Since that time, the EPA Environmental Information Management System (EIMS) has implemented the Data Catalog tables in a modified format.

(1995b). The files are maintained in a Word Perfect template and placed on the Public Web Site in plain text format.

4.4.2.4 Entries

Data Catalog entries are prepared for data sets collected and managed by EMAP. Data Catalog entries are updated when data sets are modified. Data sets that are in the Catalog include:

- early EMAP data sets;
- MAIA-Estuaries and MAIA-Surface Waters data; and
- non-EMAP data sets maintained by EMAP-IM (AED) because they have no other repository (referred to as “orphan” data sets).

The data dictionary for the Data Catalog is shown in Table 4-3.

4.4.2.5 Standards

The Data Catalog conforms to the standards established in the early EMAP program and modified in the current program (Strebel and Frithsen 1995b, EMAP 1998). The format of the Directory and Catalog are compatible with the Global Change Master Directory (GCMD) operated by NASA (NOAA 1996). The EMAP-IM system contains fields in the Data Directory and Data Catalog that allow these two components to meet FGDC standards (FGDC 1994, NOAA 1996, U.S. EPA 1996g). The compatibility of these components with GCMD and FGDC contributes to the EMAP goal of sharing information with other agencies under the CENR monitoring framework.

Section 4, Technical Design

Table 4-3. EMAP Data Catalog Fields

Data Set Identification		
<ul style="list-style-type: none"> • Title of Catalog document • Author(s) of the Catalog entry • Catalog revision date 	<ul style="list-style-type: none"> • Data set name • Task Group • Data set identification code 	<ul style="list-style-type: none"> • Version • Requested Acknowledgment
Investigator Information		
<ul style="list-style-type: none"> • Principal Investigator • Sample Collection Investigator 	<ul style="list-style-type: none"> • Sample Processing Investigator • Data Analysis Investigator 	<ul style="list-style-type: none"> • Additional Investigator
Data Set Abstract		
<ul style="list-style-type: none"> • Abstract of the Data Set 	<ul style="list-style-type: none"> • Keywords for the Data Set 	
Objectives and Introduction		
<ul style="list-style-type: none"> • Program Objective • Data Set Objective • Data Set Background Information • Summary of data set parameters • Data Acquisition • Sampling Objective 	<ul style="list-style-type: none"> • Sample Collection Methods Summary • Beginning Sampling Date • Ending Sampling Date • Sampling Platform • Sampling Equipment • Manufacturer of Sampling Equipment 	<ul style="list-style-type: none"> • Key Variables • Sampling Method Calibration • Sample Collection Quality Control • Sample Collection Method Reference • Sample Collection Method Deviations
Data Preparation and Sample Processing		
<ul style="list-style-type: none"> • Data Preparation Objective • Data Processing Methods Summary 	<ul style="list-style-type: none"> • Sampling Processing Method Calibration • Sample Processing Quality Control 	<ul style="list-style-type: none"> • Sample Processing Method Reference • Sample Processing Method Deviations
Data Manipulations		
<ul style="list-style-type: none"> • Name of New or Modified Value • Data Manipulation Description 	<ul style="list-style-type: none"> • Data Manipulation Examples • Data Manipulation Computer Code File 	<ul style="list-style-type: none"> • Data Manipulation Computer Code Language • Data Manipulation Computer Code
Description of Parameters		
<ul style="list-style-type: none"> • Parameter Name • SAS Parameter Name • Parameter label or description 	<ul style="list-style-type: none"> • Units of measurement • Parameter data type • Precision to which values are reported 	<ul style="list-style-type: none"> • Accuracy of the data values • Minimum Value in Data Set • Maximum Value in Data Set
Data Record Example		
<ul style="list-style-type: none"> • Column Names for Example Records 	<ul style="list-style-type: none"> • Example Data Records 	

Section 4, Technical Design

Table 4-3. Continued

Related Data Sets		
• Related Data Set Name	• Related Data Set Identification Code	
Geographic and Spatial Information		
• Minimum Longitude • Maximum Longitude • Maximum Latitude • Minimum Latitude	• Name of the area or region • Direct Spatial Reference Method • Horizontal Coordinate System Used • Resolution of Horizontal Coordinates	• Units for Horizontal Coordinates • Vertical Coordinate System • Resolution of Vertical Coordinates • Units for Vertical Coordinates
Data Access		
• Data Access Procedures • Data Access Restrictions • Data Access Contact Person	• Data Set Format • Information Concerning Anonymous FTP • Information Concerning Gopher	• Information Concerning World Wide Web • EMAP CD-ROM Containing the Data set
References		
• Reference Type • Reference Author • Reference Author's Affiliation • Title of Reference • Journal or Volume Title	• Journal or Volume Editor • Page and Volume Reference • Date the Reference was Published • Location of Publishing Organization	• Name of Publishing Organization • Reference Report Number or Other ID • Procite Record Number for the Reference
Quality Control/Quality Assurance		
• Measurement Quality Objectives • Quality Assurance/Control Methods • Actual Measurement Quality	• Sources of Error • Known Problems with the Data • Confidence Level/Accuracy Judgement	• Allowable Minimum Values • Allowable Maximum Values • QA Reference Data
Table of Acronyms		
• Glossary Term or Acronym	• Definition of Glossary Term or Acronym	
Personnel Information		
• Formal Title • Last Name • First Name • Middle Initial • Role • Line 1 of Address	• Line 2 of Address • Line 3 of Address • Line 4 of Address • City • State • Zip Code	• County • Voice Phone Number • Fax Phone Number • Email Address • Email Network • Additional Email Information

4.4.3 EMAP Public Web Site

4.4.3.1 Purpose

EPA information management policies lead toward sharing of data with other agencies and organizations and the public (U. S. EPA 1995a, 1995b, 1996b, 1996c). This is becoming of increasing importance with the development of the Committee on Environment and Natural Resources Framework for Environmental Monitoring. The purpose of the EMAP Public Web Site (EMAP 1998) is to provide the primary method for distributing digital EMAP data and information to all users through standard Internet browsers. Release of data and information on the site is considered a form of publication. Strebel and Frithsen (1995a) provides a summary of EMAP's commitment and EPA context for providing information through this mechanism, the issues surrounding release of information in this format, the procedures for posting information to the site, the standards and formats for the information, and the maintenance issues.

4.4.3.2 Background

The EMAP Public Web Site was implemented on the EPA Public Access Web Server at RTP in 1994 under Central EMAP-IM. Its purpose was to distribute EMAP data and information to users outside EPA. In 1996, maintenance of the site was transferred to AED, where it is now operated as part of the EMAP-IM system with guidance, review, and assistance from the IMWG.

The EMAP Public Web Site is maintained by EMAP-IM (AED) on the EPA public access server at RTP through a Working Capital Fund with the EPA Enterprise Technology Services Division (ETSD).

4.4.3.3 Design

The EMAP Public Web Site consists of a set of linked web pages that contain:

- **EMAP Program Information**—for research activities and groups;
- **EMAP-IM System Components**—Data Directory dynamically accessible through an Oracle application Web server; Data Catalog in text file format;
- **EMAP Data**—in ASCII format, Arc/Info export file, or hyperlinks to other sites where data reside;
- **EMAP publications**—publications in PDF and WP format; including the Research Strategy, data and metadata standards, MAIA Landscape Atlas, Glossary, Acronyms, Methods Format Guidance, Glossary of Quality Assurance Terms, field and laboratory operations manuals, and others;
- **Program Contacts**—lists of Working Group members; and

- **Hyperlinks**—hypertext links to related environmental programs and data sources (e.g., Global Change Master Directory, NAWQA, Chesapeake Bay Program, STORET, Envirofacts).

Preparation of these materials is a cooperative activity within EMAP. Resource Groups and Working Groups prepare data and metadata files for the site. EMAP-IM (AED) maintains the site, provides the common infrastructure, and formats some of the submitted data and information for the site. The IMWG provides guidance and review for the content and structure of the site. Data flow from the Resource Groups and Working Groups to the Public Web Site is shown in Figure 4-2.

The site contents and structure will evolve with the progress of the program (see Section 6, Implementation Plan).

4.4.3.4 Standards

Data and metadata must be prepared according to program standards and accompanied by metadata sufficient to make the data understandable and usable by unknown users over a period of years.

EMAP has developed a set of guidelines and procedures for placing data and information on the site, outlined in “Guidelines for Distributing EMAP Data and Information via the Internet” (Strebel and Frithsen 1995a) and an addendum (U.S. EPA 1998b, EMAP 1998). These guidelines specify a process “analogous to formal publication of scientific results” (Strebel and Frithsen 1995a). The process consists generally of the following four steps:

- Resource Groups and Working Groups submit data and information to EMAP-IM (AED) via FTP, email, or diskette. Materials are accompanied by appropriate publication designations, instructions, authorization, and approvals;
- EMAP-IM (AED) conducts a technical review, primarily for format; the Resource Group, Working Group, or EMAP-IM revises the submission to correct any problems found;
- EMAP-IM (AED) places the submitted information on the Internal Web Site, including placing hyperlinks and re-formatting the files for publication; the Resource Group or Working Group reviews this “proof” version for errors; and
- EMAP-IM (AED) posts the final version on the Public Web Site once approval is received from the Resource or Working Group Division Director.

Standards relevant to the Web Site also include those for preparing Data Directory and Data Catalog information, reviewed in Sections 4.4.1 and 4.4.2. Data and metadata posted on this site must have passed through stringent EMAP Project Quality Assurance Project Plan (QAPP) procedures conducted by the Resource Group or Working Group researchers. Data from projects are not always

available until 1-2 years after project completion because researchers can request that data be held until after publication of results.

4.4.4 EMAP Internal Web Site

4.4.4.1 Purpose

The purpose of the EMAP Internal Web Site is to provide a central point of access where EMAP-IM can place materials for development, testing, and review in a secured network. These materials can include data sets, information products, and data under development. The site is accessed through standard Internet browsers but is only accessible to users in the EPA Intranet domain (i.e., accessing the site from EPA computers) and does not serve the needs of non-EPA research partners.

4.4.4.2 Background

The EMAP Internal Web Site was the first networked site for EMAP data dissemination. It is maintained by AED on an ORD server in Narragansett, Rhode Island.

4.4.4.3 Design

The EMAP Internal Web Site holds the following data and tools that are not on the public Web site:

- EMAP Data Directory in Oracle, accompanied by an entry creation tool (DIF Authoring Tool), and an Oracle FORMS query tool;
- Data sets under development or being prepared for public access; and
- Draft publications, Data Catalog files, etc.

4.4.4.4 Standards

The publication process for placing data and metadata on the Internal Web Site is similar to that described for the Public Web Site (see Section 4.4.3), except that formal publication authorization is not required. The technical review for format requirements, and data and metadata completeness occur before the files are transferred to the Internal Web Site.

4.4.5 EMAP Summary Data Sets

4.4.5.1 Purpose

The process for making EMAP summary data sets available starts with the researchers who conduct collection, quality control, and analytical activities; and ends with data repositories from which data are distributed. This process is described below.

4.4.5.2 Background

EMAP data sets are in general maintained by those who collected them. Data are stored in a number of databases and formats at many different locations. However, summary data are produced for placement on the EMAP Public Web Site and distribution to all users. These data sets and their accompanying metadata are transferred from the originating organizations to EMAP-IM (AED). After review, the data and metadata are posted on the EMAP Internal Web Site for review and acceptance by the originating group and the IMWG. After appropriate review and revision, the data and metadata are moved to the EMAP Public Web Site, where all users may view and download data files, metadata, and other EMAP publications and information. The data are currently stored on the Public Web Site in ASCII and other formats.

4.4.5.3 Design

Data sets are made available as ASCII or Arc/Info export files on the EMAP Public Web Site. Data may be stored at non-EMAP web sites in other formats.

4.4.5.4 Standards

Data files must be accompanied by metadata that adequately describes the data to future users. Existing standards for quality of data managed by ORD Laboratories are useful as a guideline (U.S. EPA 1993).

4.5 System Configuration

The EMAP system configuration is based at AED and depends on AED's connections to the EPA network and EPA's Internet web servers. EPA maintains a nationwide backbone network using Cisco multi-protocol routers. Communication between Cisco sites is generally over T1 lines (see Appendix H, Configuration of the Computing Infrastructure of the Atlantic Ecology Division and National EPA). Deployment of the Cisco backbone and overall layout of the EPA network nationally and at AED are depicted in Appendix H. Operation and maintenance of the EPA backbone is the responsibility of EPA's Enterprise Technology Services Division (ETSD) at RTP. ETSD also operates and maintains any and all systems used to provide access to the public, including World Wide Web (WWW) access. EPA is committed to the principle of sharing information with other agencies and organizations and with the public. Consequently, ETSD's maintenance of EPA's networks and servers includes continual monitoring for performance issues, and upgrading resources to ensure adequate performance. RTP is using Netscape's Enterprise Server software to serve WWW information and applications, powered by the host machine, "mountain," a DEC AlphaServer 4100 with 2 CPUs running Digital Unix. The basic requirements for serving WWW information and applications are:

Section 4, Technical Design

- a suitably fast server available to the public via the Internet and capable of serving WWW applications to client browsers; and
- sufficient disk storage to contain the WWW application along with the applicable data, metadata, documents and other information.

EPA's stated intention of providing information to the public ensures that the first requirement is met. The second requirement is met because EMAP's intention is not to take custody of data, but to publish a directory of applicable data sets, so neither current disk space nor expected growth estimates present serious storage constraints.

Information management at AED utilizes a number of systems and resources available at the laboratory. The configuration of the hardware described in this section is depicted in Appendix H. A dual Pentium® Pro-180MHz processor-based Dell system serves Arc/Info coverages to workstations. Windows™ SAS applications and data are served by a Dell PowerEdge 4100/200MHz Pentium® processor-based system. A DEC Alpha 2100 Server provides Oracle™ database services while local Intranet services are being provided by the NCSA HTTPd Server software. The EMAP Directory information is maintained in Oracle tables on the Alpha server, and an Oracle Forms application is used to add and edit new directory entries. Preliminary web pages destined for the Public Web Site are also served from this system. Access to this system is limited to clients from within the EPA network only. This is useful for posting new data and metadata for review by the originating data group and ensures that no data are published before verification and approval.

Development by database programmers, information managers, and web programmers is typically done on a PC-client connected to the AED Local Area Network (Appendix H). A typical development client is a fast 486 mHz or Pentium class PC with a minimum of 32 megabytes (MB) of RAM and 500 MB or more of local disk storage. Clients at AED run under WIN95/NT as the Network Operating System. Using TCP/IP as the transport protocol allows access to all AED local services as well as remote services offered over the EPA Wide Area Network (WAN), and Internet services such as web applications.

A typical client at AED has a standard software suite based around the Windows95 environment and includes usual desktop applications such as WordPerfect, Excel, PowerPoint, Netscape Navigator, Lotus 1-2-3, and GroupWise. Users and developers with a need would also have access to Oracle development tools, SAS and ArcView. Appendix H depicts the software configurations of the AED and RTP laboratories that support EMAP-IM.

The actual storage requirements for EMAP data are relatively small at present, and are not projected to grow to unmanageable amounts. Evaluation of the storage requirements will proceed as the volumes of data collected in the current program become better known. If storage requirements change, disk storage on local AED systems can be easily upgraded using storage array building

Section 4, Technical Design

blocks. It is expected that storage needs will not approach any limiting capabilities on any of AED's servers. Likewise, AED's 4-CPU Alpha 2100 has an excess of computational capacity, and could be easily upgraded to faster CPUs should the need arise. Subsets of AED/EMAP information are routinely ported to RTP's Digital 2-CPU AlphaServer 4100 (mountain), which is configured with 1.5 GB main memory and over 400 GB of disk space. Current development clients are adequate for present demands. However, future plans will certainly require more robust systems for developers, and EMAP anticipates the need to provide upgrades for those clients.

The EPA public access server that hosts the EMAP Public Web Site can interact with any client web browser. Because the web site is configured for all standard World Wide Web access, users do not need proprietary software or hardware to browse the data or download data sets. However, use of downloaded data sets in data analysis will require that users have specialized software (e.g., GIS coverages can be downloaded with standard Web routines but the user must have GIS software to use the data).

Table 4-4. U.S. EPA (AED) Software Environment Supporting EMAP

Product	Version Level
Windows™ NT Server	4.0
DEC Unix Operating System	4.0(b)
GroupWise Mail Server	4.1(a)
Netscape™ Navigator	3.0
SFgate (gateway to WebServer)	4.0.30
freeWais/sf (search engine)	2.0.65
Perl Interpreter (CGI Script execution)	5.004
NCSA HTTPd Server	1.5.2
ArcView™	3
ARC/INFO™	7.1.2
Oracle™ RDBMS	7.3.3.5
Oracle Forms	4.5
Oracle Reports	2.5
Oracle WebServer	2.0
SQL*Net (Oracle Network Listener)	2.2
Excel	2.51 (SR-1)
Lotus 123	97
PowerPoint	7.0(b)
SAS (for Windows)	6.12 (TS020)
WordPerfect	7.0.2.26

Table 4-5. U.S. EPA (RTP) Software Environment Supporting EMAP

Product	Version Level
Windows™ NT Server	4.0
DEC Unix Operating System	4.0(b)
Verity-97 (Search Engine)	7.0
Perl Interpreter (CGI Script execution)	5.00404
Apache HTTP WebServer	1.3
Oracle™ RDBMS	7.3.3.5
Oracle Forms	4.5
Oracle Reports	2.5
Oracle WebServer	2.0
SQL*Net (Oracle Network Listener)	2.2

4.6 EMAP Archival Plan

EMAP–IM (AED) has developed a plan for long-term archival of EMAP data on their internal servers. See Appendix I, EMAP Archival Plan.

4.7 System Evaluation

The current EMAP–IM system configuration and components fulfill a number of EMAP information management requirements. This section provides an evaluation of how well the EMAP–IM system meets the requirements and what new components are needed.

4.7.1 Data Accessibility

The EMAP–IM system described in this Plan offers data accessibility at a number of levels. Individual researchers make data available for access from the EMAP system from distributed sites of their choice (including the EMAP–IM Public Web Site).

The Data Directory facilitates finding and accessing databases which are relevant to EMAP projects without the expense of capturing and maintaining the data. EMAP documentation standards promote providing users with comprehensive information for evaluating data before they are accessed. The existence of these tools does not ensure that the data and documentation will be available or delivered in a timely manner.

The Public Web Site provides access to EMAP-managed data sets in standard formats. The Internal Web Site provides a site for EMAP researchers to share data that are not ready for public release, but it is limited by a lack of capabilities for the exchange of data with non-EPA research partners. EPA policies on access to its Intranet have been articulated as a major impediment to the success of EMAP. EMAP-IM will need to add functionality to improve this capability (see Section 6, Implementation Plan).

Similar directory-based systems have been implemented by a number of national-level efforts including the Consortium for International Earth Science Information Network (CIESIN), the NASA Master Directory, and NOAA's Environmental Services Data Directory. The intent of these systems is similar to the goals stated for EMAP but these other efforts are much wider in scope. In order to increase exposure and use of the EMAP data, the EMAP Data Directory will be made accessible to these directories through Z39.50 compatibility. It would not be useful to place EMAP metadata only in these other indexes because the current system allows EMAP to manage and disseminate its own data at all the required levels, from data citations in the Data Directory to individual data sets on the Public Web Site. EMAP-IM (AED) will also monitor the development of the EPA Environmental Information Management System (EIMS), which tracks related data.

Data accessibility also includes the quality and adequate management of data sets distributed from the web sites. The EMAP-IM approach outlined in this plan ensures data quality by allowing data management to occur at the point of collection (where the expertise required to collect, QA, analyze, interpret, and maintain the data resides). Concerns of researchers about maintaining stewardship and control over their data are also addressed by this approach.

4.7.2 Flexibility of Design to Adapt to Future Technological and Program Changes

The design of the existing EMAP-IM system components are flexible enough to allow it to be scaled up to expand data management capabilities, hardware and networks, and Internet access. The only feature of the system that hinders expansion of data sharing capabilities is the limited access for non-EPA research partners to the EPA Intranet.

4.7.3 User Satisfaction

Working Groups have indicated that the implemented system adequately serves first order needs for locating and distributing their data, and they support the concept of data remaining as close as possible to the point of collection and expertise. However, they are concerned about some of the limitations discussed above, including:

- the limited tools for non-EPA researchers to exchange data in the early stages of collection and development; and
- the perceived lack of long-term commitment by a number of other data collecting organizations to managing and disseminating data they collect (EMAP may explore taking on responsibility for managing such data sets; see Section 6, Implementation Plan, and Appendix D, Preliminary Design and Options Document).

4.7.4 Benefits and Costs

The implementation of the existing system has led to enhancements of accessibility and maintenance of important data resources. The Data Directory and Public Web Site are now central EMAP resources for locating and accessing data, especially for early EMAP data and metadata.

Implementation and maintenance of the physical system resources has been relatively inexpensive, since most of the required hardware, software, and network resources already exist in EPA and can be upgraded incrementally at relatively modest cost. The most significant costs are the human resources required to assist researchers with a number of tasks, including managing data and metadata, maintaining the Data Directory and Public Web Site, distributing data sets, and maintaining data standards. EMAP currently has a significant commitment to the EMAP-IM system, and to meeting EPA goals for data sharing through interagency agreements and Internet data distribution.

4.7.5 Risks and Contingencies

As indicated above, the maintenance of this system is a human resource intensive process. EMAP-IM (AED) will need to provide review for documentation and data as updates are made. Maintenance of the Directory will require periodic reviews to ensure that automated data loading methods are working properly. Researchers knowledgeable about the projects need to prepare Directory and Catalog entries, and may need assistance from EMAP-IM (AED). EMAP-IM (AED) must ensure that Data Directory records of data sets are associated with documentation of the data sets or the data will be of limited use. EMAP-IM (AED) will be modifying the Data Directory to make it accessible to other environmental and scientific data directories (see Section 5.6).

4.8 Need for System Enhancement

The user needs, and functional and system requirements outlined in this section demonstrate the need for enhancement to the existing EMAP-IM system to serve the needs of distributed users and data. The requirements derive from the increased complexity of data flow in the post-1995 EMAP program, as well as the availability of improved technologies (e.g., World Wide Web) for data delivery. The existing system forms a robust and flexible foundation for fulfilling the basic needs of EMAP's evolving research and data integration partnerships. However, gaps in the existing

system must be addressed in order to meet the requirements outlined above for delivery and quality of data. These gaps include:

- direct access to Data Directory through Map Objects;
- compliance of Data Directory with Z39.50 bibliographic searching standard to increase the interoperability with other directory systems (e.g., GCMD);
- improved tracking of relevant GIS data in Data Directory and Public Web Site;
- additional documentation of EMAP methodology, indices, and other tools; and
- expansion of access to Internal Web Site for non-EPA research partners.

4.9 Conclusions

EPA information management policies encourage sharing of data among environmental monitoring organizations and the public (U. S. EPA 1995a, 1995b, 1996b, 1996c). Under the evolving CENR environmental monitoring framework, data directories accessible through the World Wide Web are part of the foundation for wide dissemination of data and information. The current EMAP-IM system configuration allows the EMAP-IM system to be a node in a network of national and international environmental data repository sites that index and provide access to data and information useful to EMAP users. The system components outlined here will be upgraded over the next three years to address the issues raised in 4.7, System Evaluation, and 4.8, Need for System Enhancement (see Section 6, Implementation Plan for information about future work).

Section 5

Project Management and Coordination

- 5.1 Introduction**
- 5.2 EMAP Program Management Structure**
 - 5.2.1 EMAP Management Structure**
 - 5.2.2 EMAP Program Management Related to Non-EPA Research Partners**
- 5.3 EMAP–IM Project Management Structure**
- 5.4 Information Management in Working Groups**
- 5.5 Relationships Between EMAP–IM and Information Management Authorities**
 - 5.5.1 Coordination within EPA**
 - 5.5.1.1 ORD Science Information Management Coordination Board**
 - 5.5.1.2 ORD Office of Resources Management and Administration**
 - 5.5.1.3 Coordination with EPA Office of Information Resources Management**
 - 5.5.2 Coordination with Other Federal Groups**
 - 5.5.2.1 Overview of Federal Information Management Authorities**
 - 5.5.2.2 Federal Interagency Committee on the Environment and Natural Resources**
 - 5.5.3 Coordination with Other Federal Groups**
- 5.6 EMAP–IM Project Management Challenges**
 - 5.6.1 Time Frames for EMAP Data Availability**
 - 5.6.2 Minimum Requirements for Data Delivery**
 - 5.6.3 Budgeting for Information Management**
 - 5.6.4 Ensuring that the EMAP–IM System is Fulfilling User Needs**
 - 5.6.5 Availability of Research Group Data**
 - 5.6.6 Capture of Data Sets that have No Long-Term Stewards**
 - 5.6.7 Developing Effective Relationships with other Data Repositories**
 - 5.6.8 Data Exchange Among EMAP Researchers**
 - 5.6.9 Assistance to Working Groups for Version Tracking and Documentation**
- 5.7 Conclusions**

EMAP information management involves a complex network of partnerships among groups conducting research and monitoring. The challenges of EMAP IM are to ensure adequate resources for data management by data collectors and distributors and to guarantee the long-term availability of the data to interested users. EMAP will also improve data accessibility through implementation of information standards adopted by the Federal government and cooperating research partners. These goals can be met through a combination of management and technical efforts. This section reviews the approaches being used to support effective information management in EMAP.

5.1 Introduction

In order to meet the objectives and requirements described in the EMAP–IM Plan and effectively implement the EMAP–IM system, the EMAP program must provide appropriate institutional resources and management support. EMAP is funded by a number of organizations in both true dollars and in-kind contributions. Because of the cooperative nature of the effort, EMAP is conducted in a matrix management environment in which resources are distributed to a wide array of organizations that perform the majority of the critical work. To be effective, EMAP–IM needs to develop approaches that encourage standardization and altruistic participation, as well as effective policies and operations that overcome organizational deficiencies. Existing successful information sharing examples are reviewed as models for EMAP, including the document “Promoting the National Data Infrastructure through Partnerships” (NRC 1995b). In successful programs, participating organizations share:

- responsibilities;
- commitment;
- benefits; and
- control.

The goal of the EMAP–IM Plan is to implement a system that achieves these shared objectives with the partnering organizations. Using these incentives and cooperative agreements, a solid information management environment can be constructed.

This section describes how the existing EMAP structure affects information management. Information management in EMAP is guided by the EMAP IMWG, which includes members from the ORD Laboratories (see Appendix J, Organization of ORD Offices and Laboratories), and the Working Groups who provide oversight and support to EMAP–IM. This section provides recommendations on how the rest of the EMAP participants (including EMAP management, EMAP–IM (AED), Working Groups, and research partners) can use the existing matrix management structure to support the success of EMAP data and information management.

5.2 EMAP Program Management Structure

This section describes management of the overall EMAP research program and EMAP's working relationships with non-EPA research partners.

5.2.1 EMAP Management Structure

EMAP is set up as a core program with a large number of organizationally independent research partners. The EMAP core group is responsible for:

- developing a research strategy;
- prioritizing research and monitoring projects;
- organizing pertinent information; and
- coordinating the efforts of the core program and the partners.

EMAP does not control how research partners manage and deliver data and documentation, although limited control can be exerted through cooperative agreements and grant contracts. This section addresses project management from two perspectives: First, from the perspective of the overall EMAP Program, including all partners and outside sources of data; and second, from the perspective of the core EMAP program (EMAP management and EMAP-IM).

Management and dissemination of information are more complicated in the current program than they were in the early program because EMAP managers no longer control project design, data collection, and data management. Instead, EMAP depends on a network of participating organizations to gather, manage, document, analyze, and disseminate pertinent data to its users. The success of the EMAP-IM system depends on three key perspectives driving system implementation.

- The system must meet the needs of primary users (researchers and EMAP managers) (see Section 3.2.2, Primary Users). Interactions between primary users and the IMWG provides the direction for functional and system requirements, which determines the priorities for EMAP-IM system development and enhancements.
- Data identified in the EMAP Data Directory must be available to interested users in reasonable time frames. If the data are not accessible, the system will become irrelevant to intended users.
- The EMAP-IM system must be flexible enough to serve the needs of the Working Groups, each of which has a different project management structure with different ways of completing data management and delivery tasks.

5.2.2 *EMAP Program Management Related to Non-EPA Research Partners*

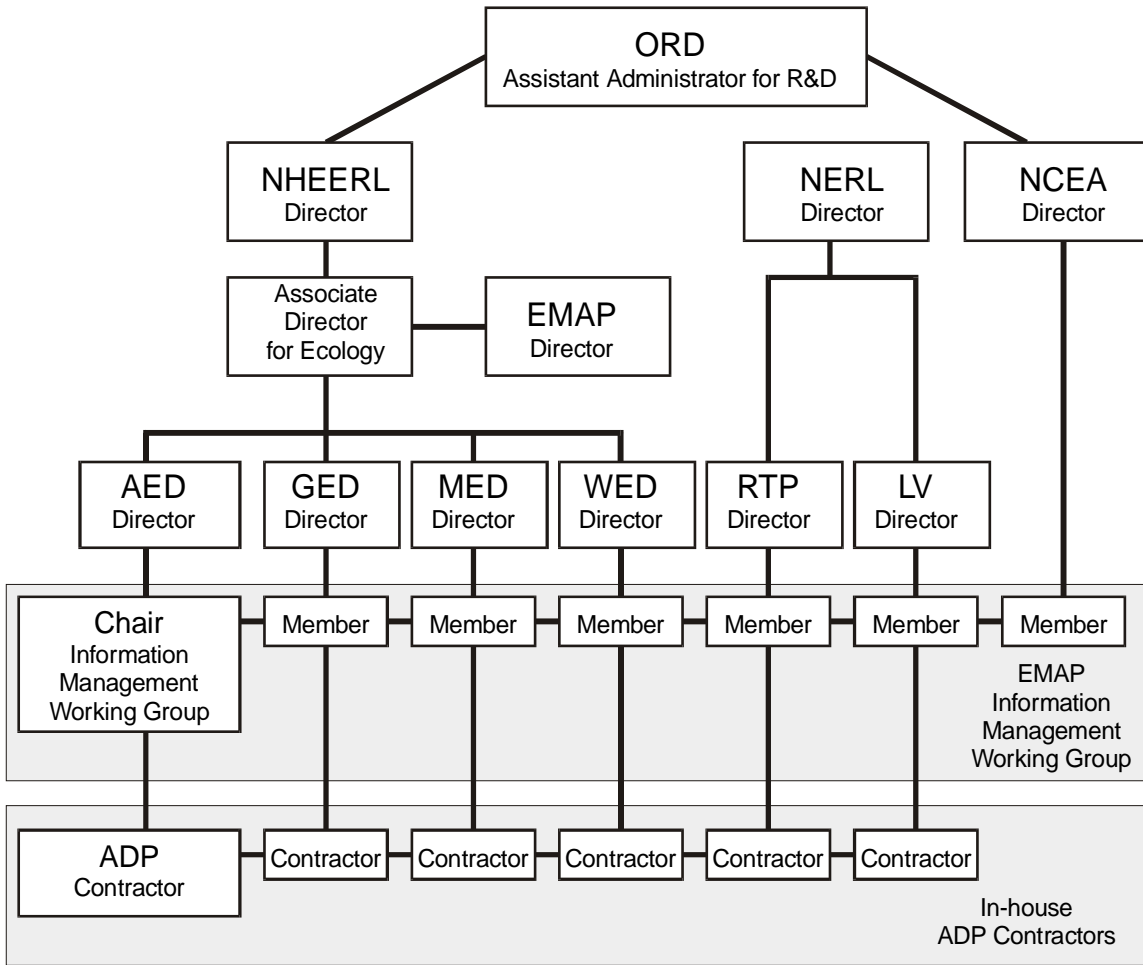
Working relationships with non-EPA research partners are governed by a number of mechanisms, including Memoranda of Understanding, contracts, federal regulations, and interagency agreements. To ensure that these cooperative research projects deliver products and data for EMAP users, it is necessary for non-EPA partners to employ standards, coordinate monitoring, and share information resources. The inter-agency agreements developed for the cooperative EMAP research projects include specifications of sampling design, data collection, analytical methodology, documentation, data management, and data analysis methodologies. They also include requirements for data submission and maintenance.

To develop interagency agreements, EMAP conducts requirements analyses and negotiations with research partners to develop agreements on IM approaches that meet the needs of all partners. Agreements are documented and revisited regularly by all parties to ensure that expectations are understood. The goal of this approach is to allow all participants to contribute to common goals that benefit them in proportion to the resources they expend. Anticipated benefits for participants include:

- reciprocal access to valuable data and information of known quality and utility from other partners;
- contributions of services from partners, such as distribution or long-term maintenance of data sets; and
- pooling of resources for research programs.

5.3 EMAP–IM Project Management Structure

The EMAP Program is led by the ORD's NHEERL. The EMAP Director reports to NHEERL's Associate Director for Ecology (Figure 5-1). Because the scope and complexity of EMAP requires extensive coordination among various laboratories and research groups within ORD, NHEERL's Associate Director for Ecology has established EMAP Working Groups with members throughout ORD.



ACN417

Figure 5-1. Structure of EMAP Organization within ORD.

Within EMAP, data and information management is led by EMAP-IM (AED), consisting of dedicated staff at AED. EMAP-IM (AED) develops and maintains the EMAP-IM system, and is advised on direction and priorities by the EMAP IMWG, which is chaired by an EPA employee from AED. The IMWG provides information standards, technical procedures, and IM guidance to EMAP. IMWG members are selected by the directors of the participating organizations. IMWG decisions are conveyed to each organization by its representative.

EMAP-IM technical information management tasks are performed on-site at AED under an agency support contract. The on-site contractor has an EMAP task leader who interacts with the AED

Information Technology Coordinator and the Chair of the EMAP IMWG for technical direction. All work done by contract employees is performed at the direction of the on-site EMAP task leader.

Many other organizations (including other ORD laboratories) involved in EMAP use similar contracting vehicles to meet technical requirements for information system design, implementation, and maintenance. The contract used at AED is an agency-wide contract which is also used by other EMAP participants. However, there is no formal coordination of contractor activities between locations. Coordination between laboratories is done by the Chair of the EMAP IMWG.

The EMAP Director influences information management by establishing program priorities and allocating funds. However, the Director has little management control over data delivery from researchers because of the matrix management approach under which EMAP is currently functioning. Actual authority over and supervision of individual research projects run by EPA offices and laboratories (including EMAP-IM (AED)) is under the control of each Laboratory Director. The Division Directors within each laboratory control technical approaches, staffing, contracts, procurement, and personnel management for resources located within their Division.

5.4 Information Management in Working Groups

Information management in the Working Groups is handled by the researchers who are located in ORD Laboratories, EPA Regions, other federal agencies, universities, research institutions, and other sites. Minimum requirements have been imposed for management of these data and information. Each Working Group is free to manage its data using its own methods in accordance with standards set by EMAP IMWG. Minimum requirements have been developed for Data Directory entries and for documentation format and content. Since EMAP-IM has no authority over these activities, they depend on the Working Groups to meet minimum requirements.

Working Groups can call on EMAP-IM (AED) for assistance with completing Data Directory entries, metadata, and long-term maintenance of data sets (see Section 5.6.9).

5.5 Relationship Between EMAP–IM and Information Management Authorities

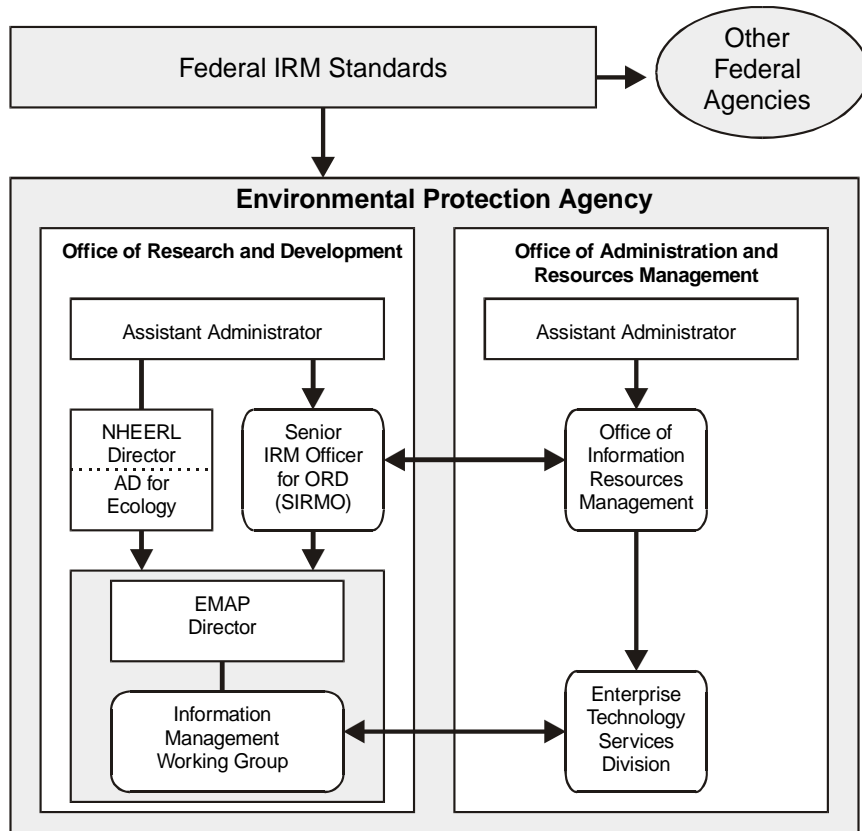
EMAP–IM must coordinate with information management activities in other relevant organizations, including Federal information authorities, ORD (U.S. EPA 1995a), EPA Headquarters and Regions, other Federal agencies, and the CENR Task Force on Data Management (TFODM). The standards, tools, and activities developed in these organizations play a role in shaping EMAP–IM policies and guidelines. The key relationships that must be maintained to coordinate development of scientific data standards are discussed in more detail in the following subsections.

5.5.1 *Coordination within EPA*

Within EPA, EMAP–IM must coordinate standards development with offices that have agency-wide authority to specify requirements and methodologies used in the implementation and maintenance of Environmental Information Systems. These groups include the ORD Science Information Management Coordination Board (SIMCorB), the ORD Office of Resources Management and Administration (ORMA), and the EPA Office of Information Resources Management (OIRM). The relationship between EMAP and these offices is depicted in Figure 5-2. EMAP–IM must also coordinate with other EPA data system development projects, such as the modernization of EPA’s Office of Water Storage and Retrieval of U.S. Waterways Parametric Data (STORET), and the development of the EPA Envirofacts (Envirofacts 1998) and the Surf Your Watersheds (SURF 1998) web sites.

5.5.1.1 ORD Science Information Management Coordination Board

Established in 1997, SIMCorB serves as a permanent body for carrying out ORD’s science information resource management (IRM) responsibilities. The Director of NERL serves as the Chairman of SIMCorB and each organizational component of ORD (including NHEERL, which leads the EMAP program) provides a representative to SIMCorB in order to ensure coordination of their information resource management efforts. The participants in SIMCorB are working to develop ORD’s 5 year Science IRM Implementation Plan and to provide ORD-wide scientific IRM policies,



ACN418

Figure 5-2. Relationship of EMAP to EPA’s OIRM and ORD’s ORMA.

procedures, and standards. Among the initial pilots for SIMCorB is the EPA EIMS, of which EMAP is an active participant. EIMS provides integrated management of data and metadata to support and facilitate environmental assessment.

5.5.1.2 ORD Office of Resources Management and Administration

Within ORD, the ORMA (Figure 5-2) Senior IRM Manager provides operational support for EPA field laboratories and central computer centers. ORMA provides support and direction to the EMAP Director concerning IRM issues. The Senior IRM Manager works with the EMAP Director to coordinate budgets and to approve major acquisitions, strategic plans, systems design, and system documents.

EMAP interactions with ORMA focus on the physical aspects of EMAP–IM systems, specifically hardware, commercial software, and networks. EMAP relies on ORMA to provide budget support and actual financial support for all EPA shared computational commercial software and telecommunications. ORMA participates actively in SIMCorB (see Section 5.5.1). Guided by the overall IRM policies and standards recommended by SIMCorB, the Senior IRM Manager will:

- provide IRM leadership;
- develop and promulgate IRM policies and guidelines;
- provide overnight, IRM reviews and technical consultation; and
- participate in discussions of new technologies, IM risks, and other relevant issues.

The Senior IRM Manager may delegate authority to a Senior IRM Officer (SIRMO) to support IRM functions in EMAP and its Working Groups. The SIRMO may delegate IRM tasks to the appropriate EMAP managers. The ORD SIRMO represents ORD’s ORMA on the IMWG to help coordinate with OIRM.

EMAP works with the ORD telecommunications group to specify network and communication requirements for EMAP–IM systems. This group supports EMAP’s access to adequate capacity and ensures that network plans are in compliance with agency standards. EMAP provides information to enable them to support EMAP telecommunication needs. It is not anticipated that EMAP will need to fundamentally enhance existing network communications beyond those already planned by the telecommunications group.

All EMAP sites have network connections (EPA Information Technology Architecture). These resources are used, in part, for the initial data analysis. In addition, many of the States cooperating in the program have provided network extensions for their specific needs. These networks are also connected to external networks such as the Internet. EMAP must provide its network requirements to the telecommunications group so it can implement, within budget limitations, the required network infrastructure.

5.5.1.3 Coordination with EPA Office of Information Resources Management

OIRM has developed an information resources strategy for EPA (U.S. EPA 1995b, U.S. EPA 1996c). The Enterprise Information Management Division (EIMD) within OIRM ensures that agency standards are met whenever possible. EIMD provides EMAP with guidelines and emerging standards for data administration in order to help ensure that agency standards are met whenever possible while still achieving EMAP objectives. EIMD provides EMAP with appropriate standards development information from such groups as National Institute of Standards and Technology (NIST), Open Systems Foundation (OSF), American National Standards Institute, Inc. (ANSI), and the International Standards Organization (ISO).

EMAP's interactions with OIRM also include:

- a Working Capital Fund agreement with OIRM's ETSD for operation of the EMAP World Wide Web Site;
- working with the OIRM telecommunications group to ensure adequate network and communication capacity for EMAP-IM systems (EMAP-IM system requirements are not expected to differ fundamentally from those currently in place or under consideration). The OIRM telecommunications group must also ensure that EMAP-IM network and communications features are in compliance with agency standards;
- improving the mechanisms for data exchange among research partners; and
- developing this IM Plan in compliance with the OIRM guidelines for EEI 1B3 (see Appendices C-G).

5.5.2 *Coordination with Other Federal Groups*

5.5.2.1 Overview of Federal Information Management Authorities

EMAP-IM operates in a context of many regulations that have been established for the creation and maintenance of Federal information systems. Major Federal Information Resource Management (IRM) guidelines that EMAP-IM must follow include: Federal Information Processing Standards (FIPS), and information management guidelines from the General Accounting Office (GAO) and Office of Management and Budget (OMB). The U.S. General Services Administration (GSA) has also defined the roles and responsibilities of Federal IRM functions as stated in the Senior Federal Information Resource Manager.

Few accepted standards exist for the management of scientific information. Hence, it is necessary to work with other agencies such as NASA, NOAA, National Science Foundation (NSF), and the U.S. Department of Energy (DOE), where significant funds are being devoted to this type of development.

5.5.2.2 Federal Interagency Committee on the Environment and Natural Resources

EMAP's participation in CENR requires coordination of research programs and data management among the CENR partners. CENR is overseeing implementation of monitoring networks among a number of Federal agencies and adoption of standards for documenting data and facilitating exchange (e.g., Z39.50, FGDC metadata standards, GCRP's GCDIS). CENR is using EMAP and MAIA as pilot programs to develop standards and policies for integrated environmental monitoring efforts, and to ensure compliance with adopted standards.

EMAP participates in CENR subcommittees that are developing standards that will shape EMAP environmental monitoring and data management efforts. EMAP IMWG members serve on the CENR TFODM in order to help determine data standards and ensure that EMAP-IM approaches are consistent with those adopted by CENR. Coordination between IMWG and CENR is currently being accomplished by those individuals who are members of both groups.

The EMAP-IM system will evolve in accordance with emerging CENR information management standards to ensure maximum interoperability of EMAP data with that of other participating agencies. Since EMAP monitoring results will be applied by a number of programs coordinated by CENR, EMAP efforts must be coordinated and reviewed within the CENR framework.

In general, CENR advocates the use of existing international standards wherever possible. EMAP has adopted this policy as well, to ensure compliance and interoperability with CENR-led initiatives for information management. The EMAP Research Plan (U.S. EPA 1997b) states that EMAP: "...will work closely with numerous CENR-led efforts to test the framework in regional-scale pilot studies." EMAP-IM will coordinate with CENR's Data Management Working Group. EMAP-IM will work towards adopting all applicable interagency data standards, guidelines, and policies to ensure maximum interoperability among relevant systems.

The CENR Information Management Subcommittee has developed a set of objectives and has begun specifying standards to help achieve them. The objectives are outlined in the document "Policy Statements on Data Management for Global Change Research" (GCRP 1998), and include:

- The U.S. GCRP requires an early and continuing commitment to the establishment, maintenance, validation, description, accessibility, and distribution of high-quality, long-term data sets;
- Full and open sharing of the suite of global data sets for all global change researchers is a fundamental objective;
- Preservation of all data needed for long-term global change research is required. For each and every global change data parameter, there should be at least one explicitly designated archive. Procedures and criteria for setting priorities for data acquisition, retention, and

Section 5, Project Management and Coordination

- purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets;
- Data archives must include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data;
 - National and international standards should be used to the greatest extent possible for media and for processing and communication of global data sets;
 - Data should be provided at the lowest possible cost to global change researchers in the interest of full and open access to data. This cost should, as a first principle, be no more than the marginal cost of filling a specific user request. Agencies should act to streamline administrative agreements for exchanging data among researchers; and
 - For those programs in which selected principal investigators have initial periods of exclusive data use, data should be made freely available as soon as they become widely useful. In each case, the funding agency should explicitly define the duration of any exclusive-use period.

EMAP has adopted these objectives in order to be consistent with CENR's decentralized data and information management approach and is working towards meeting them within the constraints of EMAP's program management structure and resources. The framework of the EMAP-IM system already follows standards and design of the GCRP GCDIS Implementation Plan (GCRI 1998). The TFODM is in the process of developing more detailed directions for information management. Until these are available EMAP will continue with its current implementation and will make changes to meet CENR requirements as needed. The current EMAP-IM system is a viable approach, and as EMAP's data distribution efforts converge with other efforts being coordinated by CENR, the current system can be made compatible with and become a data source for CENR without a major change in approach or philosophy. Modifications to the EMAP approach that will serve the CENR model are being tested in EMAP's MAIA program.

One important international standard that CENR will adopt is the ANSI/NISO Z39.50 protocol for searching and interoperability among online databases and bibliographies (Z39.50 1998). Z39.50 has been adopted by most public information sources, such as major libraries and information clearinghouses. EPA already has a cooperative agreement with one of these organizations—the Center for International Earth Science Information Network, or “CIESIN”—to develop hyperlinks from the CIESIN data clearinghouse to EPA data and information (CIESIN 1998). EMAP-IM (AED) is currently reviewing the requirements of this standard for making the EMAP Data Directory interoperable with this standard. Compliance with the standard will require support from the EPA Enterprise Technology Services Division to make the EPA public access server at RTP a Z39.50 server (see Section 6, Implementation Plan).

5.6 EMAP–IM Project Management Challenges

Because the success of the EMAP program depends strategically on the success of information management, EMAP–IM requires a strong, coherent project management plan which considers the realities of EMAP program management strategies. A number of challenges must be resolved to ensure data accessibility and the success of the research efforts. These issues are summarized in Appendix B, Data Management Needs and Practices of EMAP Working Groups).

5.6.1 Time Frames for EMAP Data Availability

Since EMAP–IM has no authority over the wide array of EMAP data sources, it will be difficult to ensure data availability. This issue was never fully resolved with the early program’s Resource Groups, who set their own data management standards and resource allocations. Time frames for data availability ranged from less than one year to more than six years. Some of the monitoring data collected in 1990 are becoming available to EMAP researchers for the first time in 1997–1998. Since the current program includes multiple non-EPA partners and even less control on the part of EMAP–IM, it is possible that these delayed time frames for data availability may persist. This problem could possibly be alleviated by specifying delivery dates for data in grants, contracts (and modifications), budgets, and interagency agreements.

5.6.2 Minimum Requirements for Data Delivery

EMAP–IM depends on data sources for data delivery and documentation, Data Directory entries, and maintenance of data sets online. EMAP–IM has developed minimum requirements for these activities but additional project management tools are needed to ensure that the standards are followed. Project management tools could include:

- standard contract language for data submission timetables and format, and data distribution;
- SOPs for the development of Data Directory entries when sampling or research begins; and
- regular EMAP–IM review of Data Directory entries and researchers’ work plans to determine whether expected data sets have not been delivered in the promised time frames (EMAP–IM can contact Principal Investigators to evaluate delays and update expected dates for data delivery).

5.6.3 Budgeting for Information Management

Budgeting issues are as important for individual EMAP programs as they are for the overall program because an individual project can cover as many disciplines and data types as larger regional and

Section 5, Project Management and Coordination

national projects. Attention to data management issues in the planning and budgeting phase could improve the accessibility of data that can be used by EMAP to aggregate the results.

Many Working Groups, however, have raised the concern that resources are not regularly allocated by EMAP or researchers for the comprehensive information management specified in this Plan. In general, work plan estimates are specified without a well-conceived and funded information management plan. Funds are often consumed prior to project completion and certainly before the data have been organized into formats optimized for interaction with the rest of the EMAP program. As a result, the groups have insufficient resources in their budget for data management, distribution, and documentation. This phenomenon is widespread in the environmental monitoring community and has been well documented by the Committee on a Systems Assessment of Marine Environmental Monitoring (National Academy of Sciences 1990):

“Data management activities are as important to the success of monitoring programs as the collection of the data. Therefore they should be funded as a continuing core program element, and reports that summarize the types, volume, and quality of the data accessible through the system should be prepared and distributed to potential users frequently. Unfortunately, monitoring data are frequently not incorporated into a data management system until most of the data collection is complete. At this point in many programs, there may not be enough time or money to create an adequate system. This situation lessens the utility of monitoring data to scientists within and outside the program.”

Data management costs can often exceed those required for initial data collection, especially if data management tasks are delayed. If sufficient resources are not allocated at the beginning of a project, data may ultimately be useful to only a few individuals. To address this shortcoming, it may be beneficial for EMAP projects to specifically allocate a minimum percentage of project budgets for information management. EMAP-IM can review all work plans to ensure that these resources have been allocated. A number of independent studies have published estimates of the minimum recommended percentage of project funds needed to properly manage data from major environmental data collection efforts. One prominent example was prepared by the U.S. GCRP. Many CENR standards are based on this program, so they will be important for EMAP. The GCRP Data and Information Management Plan (GCRP 1992) allotted 18% of its total budget to manage data in 1991, and estimated an increasing amount for subsequent years.

The National Research Council (NRC) has estimated a recommended percentage of research budgets that should be allocated to data management. In their publication “Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data; Selected Case Studies” (NRC 1995a), the Committee for a Pilot Study on Database Interfaces stated that: “While it is impossible to establish universal guidelines for funding, the committee’s investigations suggest that setting aside 10% of the total project cost for data management would not be unreasonable.” The Committee also

suggests that any environmental research should require a data management plan which demonstrates that data will be properly managed and that long-term archival is planned.

EMAP may adopt GCRP and NRC standards by ensuring that a minimum of between 10% and 18% of annual project budgets is dedicated to data management to meet whatever standards are adopted. EMAP-IM is also continuing its efforts to maintain clear, concise guidance for Working Group data management, documentation, dissemination and analysis, and directly assist researchers as needed with these tasks.

5.6.4 Ensuring that the EMAP-IM System is Fulfilling User Needs

EMAP-IM has identified regular users of the EMAP data and solicited input on the quality and availability of data sets. This group of regular users acts as the priority-setters for the EMAP-IM system. When their data needs are not being met, EMAP-IM works through the EMAP Director and the IMWG Chair to remedy the perceived deficiencies.

5.6.5 Availability of Resource Group Data

Some of the data from the early EMAP program are still not available outside of the Resource Groups that collected and analyzed it. EMAP-IM is working with the responsible Resource Groups to obtain data needed for current studies.

5.6.6 Capture of Data Sets that have No Long-Term Stewards

Some EMAP data sets useful to EMAP are collected by groups that cannot offer long-term prospects for data stewardship and distribution. These include EMAP data sets, such as data aggregates for some of the Resource Groups, and R-EMAP data sets. Many R-EMAP programs are being performed by temporary partnerships among multiple organizations. The EPA Regions coordinating them do not have consistent approaches to long-term management and dissemination of the data which reside with local and regional researchers, and have not been captured in a system that meets EMAP-IM accessibility requirements (e.g., STORET, EMAP-IM). Therefore, data from the R-EMAP programs are being summarized and delivered to EMAP-IM (AED) for long-term archival and dissemination via the EMAP World Wide Web Site. To assist with documenting and distributing these data, EMAP-IM (AED) is using available staff and computing resources at AED to capture, load, and maintain R-EMAP data on the EMAP World Wide Web Site. EMAP may request additional funding for the recovery and long-term maintenance of these data.

EMAP-IM distributes its data submission requirements to subcontractors, and Working Groups can use these standards to ensure that adequate resources and commitment language for compliance with the requirements are specified in all project work plans. Similar requirements are encouraged in all interagency agreements, Requests for Proposals (RFP's), and Memoranda of Understanding.

EMAP-IM also provides data submission guidelines to improve delivery and quality of data sets from all participants.

5.6.7 *Developing Effective Relationships with Other Data Repositories*

EMAP's future success depends on its ability to integrate data collected by others as well as becoming a source of data to non-EMAP researchers and resource managers. Many nationally and internationally recognized systems for information storage and dissemination that are similar to the EMAP-IM system already exist or are being created (e.g., CIESIN, EPA Envirofacts, Surf Your Watersheds, STORET, Safe Drinking Water Information System (SDWIS), EPA's AIRS, MDN). EMAP data stored in these systems will be cited in the EMAP Data Directory, which will provide information about the locations to allow convenient access to the data. EMAP-IM adopts standards (e.g., common formats, searching protocols, naming conventions, metadata formats, cross-referencing, communications protocols) that are compatible with other clearinghouses wherever possible. EMAP's participation in CENR's development and adoption of national and international information management standards will facilitate this goal.

5.6.8 *Data Exchange Among EMAP Researchers*

The current implementation of the EMAP-IM system allows for easy access by EPA employees to data being tested and developed on the EMAP Internal Web Site. This interchange occurs via the EPA Intranet, which allows access within EPA to preliminary data and information products so that researchers can collaborate in their development. However, Working Groups include many researchers from non-EPA organizations that do not have access to the EPA internal network (Intranet). In addition, because of security requirements established by OIRM, EMAP-IM (AED) cannot host a public FTP site. As a result, exchange and review of data under development is accomplished principally by email attachments, diskette/tapes, hard copy, or by placement of files on FTP sites hosted by other organizations. These other methods often hinder the timeliness and efficiency of data transfer and thereby do not enhance collaborative efforts in data development. A site available to all EMAP data sources is needed so that sources can view data being prepared for the EMAP World Wide Web Site, and exchange preliminary data sets with other researchers.

A number of problems hinder data exchange: email does not handle large files like those produced by Landscape Ecology, which can be hundreds of megabytes; U.S. Postal Service time frames are less than optimal for ad hoc exchange; and floppy disks and magnetic tapes are not universally compatible among the computer systems of all research groups. To meet this requirement, EMAP-IM (AED) is working with OIRM and SIMCorB to establish solutions to this problem, which has repeatedly been identified as a major issue. It is likely that EPA's ETSD at RTP will develop an Extranet capability which would allow approved non-EPA users access to the Internal Web Site (which is run on secured EPA servers) (see Section 6.2.4, EMAP Internal Web Site). This capability would promote effective data sharing and enhance incentives for non-EPA research partners to participate in effective, timely data exchanges.

5.6.9 Assistance to Working Groups for Version Tracking and Documentation

EMAP–IM provides guidance and assistance to Working Groups in a number of areas, since the information management expertise in these groups varies. Areas of importance include:

- Database management tools. Many of the data management tools now being used in Working Groups are sufficient for technical personnel to manage data for their own research purposes, but are not optimal at allowing researchers to track data availability, content, status, and updates. Managing multiple versions of data sets is a major data management and documentation issue that may require application of standard version control software and expertise currently not available in most Working Groups (see Section 4, Technical Design);
- Distribution of data, procedures, and results. Management of the data is required but not sufficient to make the data useful. In addition, procedures, documentation, and summary of results obtained by trained scientists is necessary. Some organizations related to EMAP may need assistance with distribution of data and documentation, procedural codes, and summary results; and
- Development of acceptable documentation compliant with CENR, EMAP and FGDC standards. Documentation should be made available in standard Federal or EMAP formats in order to ensure that other programs can successfully incorporate EMAP data. The FGDC documentation format fits well into the EMAP Data Directory and Data Catalog model, and FGDC documentation files can be linked or converted to the Data Directory and Data Catalog.

5.7 Conclusions

The success of EMAP's decentralized, interagency program management structure relies on mutual interest of participants and the positive effects of cooperation and coordination. EPA ORD and the EMAP Director have only a minimal amount of actual management control over information management in the form of negotiated interagency agreements and funding leverage. Primary responsibility for data management and information dissemination remains the domain of the data sources and organizations leading the monitoring efforts because they are best able to maintain the long-term integrity of the data. EMAP-IM's role is to index the available data sets in a directory and maintain data sets that have no other long-term stewards.

In order for the interagency agreements to be effective, standards for data delivery and documentation need to be included. These standards can require that a minimum percentage of research project budgets be set aside for information management (including development of project data management plans, data submission, and preparation of Data Directory entries and FGDC-compliant metadata). Adoption of minimum standards is consistent with the practices of other leading national environmental information systems efforts (GCRP 1998).

In order to address these priorities, EMAP is focusing on the following information management activities and policies:

- specifying minimum information management standards and budgets in all interagency agreements;
- keeping track of expected data sets in a status table to track progress of anticipated data products;
- working closely with other large scale information management and dissemination efforts, within EPA and other federal agencies;
- encouraging EMAP associates, principal investigators, and partners to store data in existing systems with demonstrated longevity and success in managing, maintaining, and disseminating data; and
- participating actively in standards development with CENR and adopting existing relevant standards (especially those of CENR and FGDC).

These factors are critical to the success of EMAP information management, and to fulfilling the mission and goals of the EMAP program.

Section 6 Implementation Plan

- 6.1 Introduction**
- 6.2 System Components—Maintenance and Enhancement**
 - 6.2.1 Data Directory**
 - 6.2.1.1 Management and Coordination**
 - 6.2.1.2 Tasks**
 - 6.2.1.3 Responsibilities**
 - 6.2.1.4 Resources Needed**
 - 6.2.2 EMAP Data Catalog**
 - 6.2.2.1 Management and Coordination**
 - 6.2.2.2 Tasks**
 - 6.2.2.3 Responsibilities**
 - 6.2.2.4 Resources Needed**
 - 6.2.3 EMAP World Wide Web Site**
 - 6.2.3.1 Management and Coordination**
 - 6.2.3.2 Tasks**
 - 6.2.3.3 Responsibilities**
 - 6.2.3.4 Resources Needed**
 - 6.2.4 EMAP Internal Web Site**
 - 6.2.4.1 Management and Coordination**
 - 6.2.4.2 Tasks**
 - 6.2.4.3 Responsibilities**
 - 6.2.4.4 Resources Needed**
- 6.3 Early EMAP (1990–1995) Data**
 - 6.3.1 Management and Coordination**
 - 6.3.2 Tasks**
 - 6.3.3 Responsibilities**
 - 6.3.4 Resources Needed**
- 6.4 Current EMAP (1996–) Data**
 - 6.4.1 Management and Coordination**

Section 6, Implementation Plan

- 6.4.2 Tasks**
 - 6.4.3 Responsibilities**
 - 6.4.4 Resources Needed**
- 6.5 Data Management**
 - 6.5.1 Management and Coordination**
 - 6.5.2 Tasks**
 - 6.5.3 Responsibilities**
 - 6.5.4 Resources Needed**
- 6.6 GIS Spatial Data and Analysis**
 - 6.6.1 Management and Coordination**
 - 6.6.2 Tasks**
 - 6.6.3 Responsibilities**
 - 6.6.4 Resources Needed**
- 6.7 Mid-Atlantic Integrated Assessment**
 - 6.7.1 Management and Coordination**
 - 6.7.2 Tasks**
 - 6.7.3 Responsibilities**
 - 6.7.4 Resources Needed**
- 6.8 Western Pilot Regional Assessment**
 - 6.8.1 Management and Coordination**
 - 6.8.2 Tasks**
 - 6.8.3 Responsibilities**
 - 6.8.4 Resources Needed**
- 6.9 EMAP–IM System Administration and Coordination**
 - 6.9.1 Tasks**
- 6.10 Overall Resource Requirements for EMAP–IM (AED)**
- 6.11 Conclusions**
- 6.12 Tentative Schedule**
 - 6.12.1 FY1999 Tasks**
 - 6.12.2 FY2000 Tasks**
 - 6.12.3 FY2001 Tasks**

The EMAP–IM system will be enhanced over the next three years (1998–2001) to improve its functionality for delivering data and information to users. This section provides a detailed schedule of the tasks that will be implemented to continue developing system components, data management capabilities, and working relationships among partners.

6.1 Introduction

This section describes the maintenance, enhancement activities, and resources that will be implemented over the next three years to continue development of the EMAP Information Management system. The tasks outlined below are designed to meet the requirements expressed by the Working Groups (see Section 3, Information Management Needs and Requirements, and Appendix B, Data Management Needs and Practices of EMAP Working Groups) and address some of the project management issues raised in Section 5 (Project Management and Coordination). The tasks will be implemented in the stages outlined in Section 6.12, Tentative Schedule. Two principal efforts will be to complete documentation of EMAP data sets in the Data Directory and enhance the functionality of the Directory and the EMAP World Wide Web Site. EMAP-IM (AED) coordinates all modifications with the EMAP IMWG, who review and approve changes before they are implemented. Final decisions are made by the IMWG Chair.

6.2 System Components—Maintenance and Enhancement

Enhancement of the EMAP-IM system components is described in this section.

6.2.1 Data Directory

As the core of the EMAP-IM strategy, maintenance and enhancement of the Data Directory is a high priority.

6.2.1.1 Management and Coordination

The main objectives are to ensure that all EMAP-funded data sets are documented in the Data Directory and to improve Internet access to the Directory by making it accessible through Web Server tools.

6.2.1.2 Tasks

The tasks that will be undertaken to enhance the EMAP Data Directory are explained in this section, with an indication of the groups responsible and the management and coordination needs.

Maintain and update existing entries as necessary, make new entries for emerging data sets
EMAP-IM (AED) has taken the lead for populating the Data Directory with entries and for assisting researchers with preparing the entries. Entries for Resource Group data have been ongoing since 1991, and are now being completed. For Resource Groups no longer in existence, such as Agroecosystems, entries are being made by the EMAP-IM (AED) ADP Contractor. EMAP-IM (AED) is also assisting Working Groups with this task.

Section 6, Implementation Plan

Maintain Data Directory Oracle database on EMAP World Wide Web Sites

The database allows direct user access to the Oracle database, thereby making the Directory more accessible, increasing search capabilities, and eliminating redundancy and version problems. It also allows the storage of hyperlinks to data repository sites directly from the Directory.

Review modified EIMS Data Directory/Catalog

EMAP-IM will review the Environmental Information Management System being implemented by NCEA to evaluate its potential functionality for EMAP and other related monitoring and research programs.

Transfer EMAP Data Directory entries to NCEA for addition to EIMS

Entries from the EMAP directory will be downloaded to EIMS for inclusion (as a duplicate set) in that database.

Request that EPA add Z39.50 protocol to RTP public access server

EMAP-IM will investigate modification of the Data Directory to be compliant with the ANSI/NISO Z39.50 information retrieval standard (Z39.50 1998) in order to comply with information delivery standards being considered by CENR. This revision will facilitate two-way exchange of directory information between EMAP and other major environmental research and monitoring programs. To accomplish this conversion, EPA will need to make the RTP server Z39.50-compliant.

Revision of Oracle database

The Oracle database will be updated for compliance with metadata standards and possibly for Z39.50 compliance.

Revision of Oracle Client

The existing Oracle client for maintaining the Data Directory database will be updated.

Update Data Directory Policy, Guidelines, and Standards Manual

EMAP-IM will review, modify and update policies, guidelines and standards for developing and maintaining the Data Directory and its entries. This activity will be ongoing with milestones in February 1999, 2000, and 2001.

Synchronize with U.S. Global Change Research Program Data Directories

CENR is adopting GCRP Data Directory approaches and standards. EMAP will make data available in formats compatible with the GCRP, either by adopting Z39.50 protocol or placing data in GCRP-accessible databases that have Z39.50 capability.

Section 6, Implementation Plan

Revise Directory to be Compatible with Z39.50 protocol

The Data Directory will be converted to the ANSI/NISO Z39.50 standard (see Section 5.5.2.2, Coordination with Federal Interagency Committee on the Environment and Natural Resources).

Load Western Pilot Directory entries

Data Directory entries will be loaded for Western Pilot data sets.

6.2.1.3 Responsibilities

The responsibilities of EMAP–IM (AED) and Working Group researchers for ensuring completeness of entries in the Data Directory are summarized below.

EMAP–IM (AED):

- maintain existing Directory to ensure the flow of information to users;
- lead maintenance of Directory database in Oracle;
- perform modifications and enhancements to Directory as opportunities and requirements arise (availability of resources, new program requirements, and new technologies);
- support Resource Groups and Working Groups with development of Directory entries by assisting with entry creation; providing standards and Directory entry creation tools; conducting quality assurance; loading and maintaining entries;
- coordinate transfer of Data Directory entries from researchers;
- develop Directory entries for data sets with no guaranteed long-term stewardship (e.g., 1990–1995 Estuaries, Great Lakes);
- work towards public web browser access to the Directory Oracle database;
- evaluate application of emerging standards to Data Directory by coordinating with organizations that develop standards; and
- coordinate ability of EMAP Data Directory to be searched by other organizations according to protocols and standards developed in the Global Change Research Program.

Resource Groups and Working Groups:

- create Data Directory entries for all data sets; and
- assign Principal Investigator or appointed contact person to coordinate with EMAP–IM on delivery of Directory entries according to EMAP standards for Directory format and placement of material on EMAP World Wide Web Site (U.S. EPA 1998b).

Section 6, Implementation Plan

6.2.1.4 Resources Needed

The continued maintenance and enhancement of the EMAP Directory requires commitment of adequate personnel in Resource Groups and Working Groups who create Directory entries, at AED where the Directory Oracle database is maintained. This activity requires allocation of resources within projects at the levels discussed in Section 5.6.3 (Budgeting for Information Management). Hardware and software requirements are the same as those of the labs where data are maintained and the server capabilities at RTP for the Web Site, and will not be specified in this Plan. Hardware and software requirements for EMAP-IM (AED) are discussed in Section 3.5, System Requirements.

6.2.2 *EMAP Data Catalog*

The EMAP Data Catalog provides a standard EMAP format for documenting data sets. In this section, enhancements will be described that will make the documentation more accessible to users and increase its compliance with emerging Federal standards.

6.2.2.1 Management and Coordination

EMAP-IM (AED) coordinates maintenance and enhancement of the Catalog as text files in the EMAP-IM system. All Resource Groups are responsible for completing Catalog entries of available 1990-1995 data sets. Each former Resource Group should ensure completion of metadata files (described in Section 6.2.1, Data Directory, above).

EMAP-IM (AED) will assist all data originators with completing the entries. EMAP-IM (AED) will supply the Resource Group information manager with EMAP Data Catalog standards and will help on an as-needed basis with development and processing of the metadata files. EMAP-IM (AED) also conducts quality assurance testing on submitted files according to EMAP standards (Frithsen and Strebel 1995; Strebel and Frithsen 1995a, 1995b; U.S. EPA 1996g, 1996h, 1998b). When files have been completed and cleared through the review process, EMAP-IM (AED) loads them onto the Web Site.

6.2.2.2 Tasks

The following tasks are planned for EMAP Data Catalog enhancement and maintenance.

Migrate from WordPerfect template to specialized Metadata format

EMAP-IM will convert the Data Catalog to a more standardized database format like those used by other organizations (e.g., FGDC) for creation and maintenance of metadata. It is likely that in the next few years, EMAP will adopt FGDC's MetaMaker (NBII 1998), which is currently being extended by the Chesapeake Bay Program (CBP). This tool will allow the storage of EMAP data in a widely accepted standard format that will be more easily accessible to the Internet. Use of such a standard format may make it possible to use metadata entry tools that have already been developed for this metadata tool.

Section 6, Implementation Plan

Update Data Catalog Policy, Guidelines, and Standards Manual

EMAP–IM reviews, modifies and updates policies, guidelines and standards for developing, maintaining and disseminating metadata on an ongoing basis. Milestones for updating these areas are scheduled for February of 1999, 2000, and 2001.

6.2.2.3 Responsibilities

The responsibilities of Resource Groups and EMAP–IM (AED) for ensuring completeness of Data Catalog files are summarized below.

EMAP–IM (AED):

- maintain and update existing Catalog files to ensure the availability of metadata to potential data users;
- support Resource Groups by assisting with development of Catalog files, providing standards and creation/formatting tools;
- coordinate transfer of Catalog files from researchers;
- conduct review of submitted files;
- develop Catalog files for data sets with no guaranteed long-term stewardship;
- load and maintain entries on Web Site;
- convert Catalog format to a more widely accepted standard to increase its accessibility and usefulness for Web searching; and
- maintain the compliance of the Catalog format with Federal standards, and evaluate application of these standards by coordinating with organizations that develop standards.

Resource Groups and others using this format:

- create Catalog files for all data sets; and
- ensure that information management contact with EMAP–IM (AED) coordinates delivery of Catalog files according to EMAP standards for Catalog and placement of material on EMAP World Wide Web Site (EMAP 1998).

6.2.2.4 Resources Needed

The EMAP Data Catalog requires adequate commitment of project personnel (researchers and EMAP–IM) and time to documenting EMAP data sets in a format that is useful to future users and accessible to Web Browser searching. It may also require purchase of metadata software (e.g.,

Section 6, Implementation Plan

MetaMaker). Creation of metadata requires allocation of resources within projects at the levels discussed in Section 5.6.3 (Budgeting for Information Management).

6.2.3 EMAP World Wide Web Site

The EMAP World Wide Web Site is one of the most important components in the EMAP–IM data access strategy, since it provides a gateway to data managed at many distributed sites. It is also currently the main tool for managing EMAP summary data sets. It contains Working Group web pages, and will soon house the EMAP Bibliographic Database. The Web Site will evolve throughout the program. This section outlines the implementation issues and tasks that will be addressed over the next three years.

6.2.3.1 Management and Coordination

EMAP–IM (AED) maintains the EMAP World Wide Web Site under a Working Capital Fund agreement with the EPA OIRM Enterprise Technology Services Division (ETSD) in RTP (see Section 6.2.3.5, Working Capital Fund Agreement). The Web Site content and format is overseen by the IMWG, and EMAP participating organizations have input through their IMWG representative. Resource Groups and Working Groups prepare their information (data sets, web pages, metadata) for the web site. EMAP–IM (AED) supports their efforts by providing standards for placement of information on the Site, provides assistance for developing the web pages, creates hyperlinks pointing to EMAP data at other sites, and implements the web pages on the site.

6.2.3.2 Tasks

The EMAP World Wide Web Site is being upgraded to provide complete coverage of EMAP data and metadata, and to add web server capabilities for the Data Directory and for GIS data in order to improve the search capabilities, delivery options, and management of the information. The specific tasks and organizations responsible are outlined in this section. The maintenance and enhancement of the EMAP World Wide Web Site is an ongoing effort and will continue as long as the EMAP program is active. The major tasks that will be carried out during 1998–2001 to achieve these goals are outlined below.

Ongoing maintenance and expansion of content

EMAP–IM (AED) will continue to expand its base of information about data useful to EMAP researchers. EMAP–IM will explore use of the EPA Spatial Data Library system and other EPA sites that track GIS data, in order to provide additional access to GIS data. EMAP–IM (AED) will also assist Working Groups with creating web pages, as needed. The EMAP Bibliographic database will be added to the Web Site.

Section 6, Implementation Plan

Major redesign to incorporate more database capabilities

The web site redesign will enhance delivery of information from the relational databases developed by EMAP-IM (AED) and make them accessible using standard web browsers from the Public Access Web Site.

Request that ETSD add MapObjects capability to EPA Public Access Web Site and develop this capability on the EMAP World Wide Web Site

EMAP-IM has requested that the RTP public access server add functionality allowing users to access available EMAP georeferenced data from GIS and remote sensing research in map formats. This functionality can be provided by installation of a MapObjects (ESRI) server license on the RTP public access server. This functionality will be added to the EMAP World Wide Web Site after it has been fully tested on the EMAP Internal Web Site.

Review of Internal and Public Access Web Sites by EMAP IMWG periodic reviews (ongoing)

The EMAP IMWG will review the content and structure of the web sites in January of 1999 and 2000.

Update Web Publishing Policy, Guidelines, and Standards Manual

EMAP-IM will review, modify and update policies, guidelines and standards for developing, maintaining and disseminating data via the EMAP Web Site. These activities will be conducted on an ongoing basis with milestones in February of 1999, 2000, and 2001.

Add EMAP Bibliographic Database to the EMAP World Wide Web Site

EMAP-IM has created a searchable bibliographic database of EMAP publications (journal articles, technical reports, planning documents, other literature) in Oracle that will be placed on the Public Access Web Site for searching and access by all users. The specific tasks include:

- Load cleaned data, revise Web-based query form, write data submission format guidelines (develop policies, guidelines and standards for developing, submitting and maintaining bibliographic entries on the EMAP public access site); move database from internal web site to public access web site;
- Load Western pilot bibliography, update data submission format guidelines; and
- Load new entries.

Maintain information about data status

EMAP-IM maintains status tables for EMAP information on the Public Access Web Site that indicates the status of Data Directory entries, data files, and metadata files. Some of the data sets noted in the status table may never be posted on the Web Site or entered into the Directory, but they are noted in the table for completeness of program documentation.

Section 6, Implementation Plan

6.2.3.3 Responsibilities

The responsibility for the content and functioning of the EMAP World Wide Web Site is shared by EMAP-IM (AED), the Resource Groups and Working Groups, and ETSD.

EMAP-IM (AED):

- maintains the web site and implements new pages;
- creates web pages with EMAP content;
- coordinates with Resource Groups and Working Groups to incorporate their data and metadata or prepare hyperlinks to sites where the data and metadata reside;
- transfers all data and metadata to web site;
- evaluates future requirements and emerging technologies to ensure that the web site is optimized to deliver information to users; and
- works with ETSD to evaluate options for access policies and security solutions.

Resource Groups and Working Groups contribute material for input to web pages in accordance with the EMAP standards for placing material on the Web Site (U.S. EPA 1998b) on the projects they lead:

- Estuaries Virginian Province and Carolinian Province: Atlantic Ecology Division;
- Estuaries Louisianian Province and West Indian Province: Gulf Ecology Division;
- Great Lakes: Mid-Continent Ecology Division;
- Surface Waters and Wetlands: Western Ecology Division;
- Forests, Rangelands data; MAIA GRD: NERL, Las Vegas;
- Landscape Characterization: NERL;
- Agro-ecosystems: U.S. Department of Agriculture;
- ORD Regional Assessments/MAIA—Overall coordination, Atlantic Ecology Division
 - Estuaries—Atlantic Ecology Division, Chesapeake Bay Program, NOAA
 - Surface Waters—Western Ecology Division
 - Landscape Ecology—NERL/Las Vegas;
- Intensive Sites—Overall coordination, Gulf Ecology Division

Section 6, Implementation Plan

- DISPro—NERL/NCEA, National Park Service, Individual researchers, NUVMC, Western Ecology Division
- CISNet—Gulf Ecology Division, NOAA, NASA
- Aquatic Mortality Database—Atlantic Ecology Division, Gulf Ecology Division;
- Landscape Ecology—NERL/Las Vegas;
- R-EMAP—(Overall coordination: Mid-Continent Ecology Division) EPA Regions, Individual researchers (states, universities, etc.); and
- Ecological Indicators—Gulf Ecology Division.

The EPA Enterprise Technology Services Division, through a Working Capital Fund Agreement with EMAP, contributes to the efficient running of the system in order to deliver access to the data needed by EMAP users. ETSD provides the following services:

- maintains the Web Site and links to the EPA Web Site;
- enhances functionality of Directory and Web Site by improving search capabilities (Web Server tools) and compliance with standards (e.g., Z39.50); and
- provides technical assistance to AED on new technology (e.g., Web Server tools) to enhance functionality of the Web Site and its components.

6.2.3.4 Resources Needed

Maintaining and adding content to the Web Site will require sufficient resources allocated within EMAP-IM (AED), Resource Groups, Working Groups, and at ETSD.

- Personnel: EMAP-IM (AED), Resource Groups, Working Groups, and ETSD will need to allocate adequate personnel to:
 - EMAP-IM (AED): implement and maintain web pages, quality assure materials to be placed on the web site, coordinate materials with data sources, coordinate upkeep and format of web site with ETSD;
 - Resource Groups and Working Groups: create data, metadata, documents, web pages and submit to EMAP-IM (AED); and
 - ETSD: Maintain site on RTP public access server and links.
- Hardware:

Section 6, Implementation Plan

- ETSD: existing hardware at RTP with routine upgrades is sufficient at present for hardware, software, network services, security, and personnel. Distribution of data via Oracle database may require additional hardware.
- Software:
 - EMAP-IM (AED): purchase Web editing tools; and
 - ETSD: use existing software at RTP and add Oracle web server software.

6.2.4 EMAP Internal Web Site

The Internal Web Site provides an important function for EMAP as the site for testing and evaluation of EMAP data and information products that are under development. The Oracle version of the Data Directory is maintained here with a web server tool that allows online searching and the DIF authoring tool for creating Directory entries. Data, metadata, and other information products are placed here for final quality assurance before being placed on the Public Access Web Site.

6.2.4.1 Management and Coordination

EMAP-IM (AED) maintains the Internal Web Site on a server at the Atlantic Ecology Division that is connected to the EPA internal network. All EMAP participants that are part of EPA can access the site, although non-EPA partners cannot. The inaccessibility of the site to non-EPA research partners is being evaluated and solutions are being developed.

6.2.4.2 Tasks

The primary tasks will be to test new functionality for the Public Access Web site that will be designed to serve data, information, and graphic files (e.g., maps) to the Internet; and to improve access to the site for non-EPA EMAP research partners.

Add ESRI MapObjects capability to Internal Web Site

EMAP-IM (AED) will develop and test the ESRI MapObjects application for providing users with access through web browsers to available georeferenced (map) data that has been created using GIS and remote sensing applications.

Add capability for partners outside EPA to access preliminary data

EMAP will explore options for broadening access for non-EPA EMAP partners to data prior to the completion of Quality Assurance. EMAP currently can only exchange such data sets through email, external FTP sites, and magnetic media (disks). For EMAP research partners that need to exchange data, expanded access (e.g., extranet) to the EMAP Internal Web Site will be addressed by EMAP-IM (AED) and RTP in the future (see Section 5.6.8, Data Exchange among EMAP Researchers).

Section 6, Implementation Plan

6.2.4.3 Responsibilities

EMAP-IM (AED) has primary responsibility for maintenance of the site. Data sources such as Resource Groups and Working Group researchers are responsible for providing material for the site. AED ADP staff and contractors maintain the server and internal network that host the site.

6.2.4.4 Resources Needed

Existing resources at AED (software, hardware, personnel) are sufficient to maintain the site in its current configuration. To implement the map server tool, it will be necessary to purchase the appropriate software. To implement the improved access for non-EPA research partners will require the cooperation of OIRM to determine new configurations for allowing EPA data to be shared with non-EPA research partners involved in cooperative agreements.

6.3 Early EMAP (1990–1995) Data

6.3.1 *Management and Coordination*

EMAP-IM (AED) is continuing its efforts to make quality-assured summary data and metadata from the Resource Groups available on the EMAP World Wide Web Site. The EMAP-IM system provides a structure for Resource Groups to share their data. EMAP-IM (AED) can directly support the Resource Groups with data and metadata preparation. Resource Groups prepare the summary data and metadata, and will supply updates as needed.

Each ORD Lab responsible for an EMAP Resource Group has an information management contact who coordinates transfer of data and metadata to EMAP-IM (AED) for posting on the Web Site. EMAP-IM (AED) supports this person by providing guidelines for posting data and metadata on the Web Site, assisting with preparation of metadata and Data Directory entries, and transferring data sets and metadata to the Web Site.

6.3.2 *Tasks*

The basic EMAP activities planned at this time for Resource Group data sets are maintaining the status table of incoming data sets, completing metadata, and transferring data files, metadata, and GIS coverages to the EMAP World Wide Web Site. Specific tasks planned by EMAP-IM (AED) are outlined in the following subsections.

Finish Remaining Resource Groups Data

EMAP-IM will document, load onto the Web Site and cross-reference the data from the Forests and Agroecosystems groups when these data are made available.

6.3.3 Responsibilities

The following groups have responsibility for long-term management, summarization, and documentation of EMAP Resource Group data:

- Atlantic Ecology Division: Estuaries Virginian Province and Carolinian Province data;
- Gulf Ecology Division: Estuaries Louisianian Province and West Indian Province data;
- Mid-Continent Ecology Division: Great Lakes data;
- Western Ecology Division: Surface Waters and Wetlands data;
- NERL, Las Vegas: Forests, Rangelands data; MAIA GRD;
- NERL, RTP: Landscape Characterization data; and
- U.S. Department of Agriculture: Agro-Ecosystems data (Directory entries and metadata will be developed by USDA under Interagency Agreement (IAG)).

6.3.4 Resources Needed

The former EMAP Resource Groups must be given adequate information management personnel to complete data documentation and transfer as described above. These will not be trivial tasks, especially because of the loss of institutional knowledge for information management that occurred when the program was re-directed in 1995.

6.4 Current EMAP (1996–) Data

One of EMAP–IM’s most significant tasks will be to ensure the availability and documentation of data produced in the Working Groups. This section describes the tasks, management, and resources needed to complete this task.

6.4.1 Management and Coordination

The success of the decentralized model for managing EMAP data and metadata will depend on tracking and cross-referencing sites where EMAP data are maintained and documented. The EMAP Data Directory and Web Site are already being used for this purpose, and many EMAP data sources are referenced and linked on EMAP or EPA web sites (e.g., The Southern California Coastal Water Research Project Authority (SCCWRP), Multi-resolution Land Characteristics Consortium (MRLC), Chesapeake Bay Program, and the Great Lakes Program).

To ensure delivery of EMAP data and metadata from decentralized sites, EMAP–IM (AED) will coordinate with researchers and participants in Working Groups to create the necessary Data Directory entries, and hyperlinks for all EMAP data. EMAP–IM will also coordinate with data

Section 6, Implementation Plan

sources to support creation of metadata. All efforts will be made to get the data sources to develop the documentation themselves; in some cases, EMAP-IM (AED) may develop the documentation and request review and comments from the data originator. EMAP-IM (AED) also tracks the status of data sets in progress by maintaining a status table on the Public Access Web Site. EMAP-IM (AED) will assume responsibility for developing Directory entries and metadata from programs that no longer have resources to support data management.

To improve the quality of data and metadata available to future EMAP users, EMAP-IM (AED) will maintain standards for submission of Data Directory entries and metadata in order to encourage compliance with FGDC standards and adequate information for future EMAP users. The early and current EMAP-IM efforts have established a number of useful information management standards that are now being successfully used by data originators such as Estuaries and Surface Waters. As information management standards are developed and adopted by other Federal programs like CENR and GCRP, EMAP will adopt a conservative approach of retaining its established standards until new standards are proven useful and have been adopted by groups with which EMAP works closely (e.g., GCRP, CENR). As EMAP adopts new standards, it may still allow use of earlier EMAP standards that may be modified to comply with the new standards. For example, EMAP is still using the EMAP Data Directory standard, but has added fields to the tables to comply with FGDC metadata standards.

6.4.2 Tasks

The following tasks are planned to meet the requirement for availability of EMAP data and metadata.

Load 1996–2001 Data and Metadata onto EMAP World Wide Web Site

EMAP-IM will load data onto the EMAP World Wide Web Site, including:

- ORD Regional Assessments/MAIA
 - Estuaries—1997 and 1998 data
 - Surface Waters—1997 and 1998 data
 - Landscape Ecology—landscape indicator coverages
 - Western pilot—load remaining data
- Intensive Sites
 - DISPro/UV-B data—finish database design and web site design, expand to sites in Western pilot area and load FY2000 data, load FY01 data

Section 6, Implementation Plan

- ❑ R-EMAP—complete the documentation, loading and cross-referencing of available R-EMAP data sets. Load data from FY1998 studies, FY1999 studies, and FY2000 studies.

6.4.3 Responsibilities

The following organizations share responsibility for maintaining and documenting 1990–1996 Working Group data:

- ORD Regional Assessments/MAIA—Overall coordination Atlantic Ecology Division
 - ❑ Estuaries—Atlantic Ecology Division, Chesapeake Bay Program, NOAA
 - ❑ Surface Waters—Western Ecology Division
 - ❑ Landscape Ecology—NERL/Las Vegas;
- Intensive Sites—Overall coordination Gulf Ecology Division
 - ❑ DISPro—NERL/NCEA, National Park Service, Individual researchers, NUVMC
 - ❑ CISNet—Gulf Ecology Division, NOAA, NASA
 - ❑ Aquatic Mortality Database—Atlantic Ecology Division, Gulf Ecology Division;
- Landscape Ecology—NERL/Las Vegas;
- R-EMAP—(Overall coordination: Mid-Continent Ecology Division) EPA Regions, Individual researchers (states, universities, etc.); and
- Ecological Indicators—Gulf Ecology Division.

These Working Groups need to identify at least one person as the key contact with EMAP–IM. At a minimum, each Working Group should be allocated sufficient resources to have a database manager and a data librarian to organize data and develop concise, comprehensive documentation. One of these staff members can also be the key contact to EMAP–IM.

6.4.4 Resources Needed

It is important that Working Groups and their researchers allocate sufficient resources for maintenance and documentation of data. It is recommended that a minimum percentage of project budgets should be allocated for information management, and EMAP–IM can review all work plans to ensure that sufficient resources have been allocated (see Section 5.6.3, Budgeting for Information Management, and Section 5, Project Management and Coordination).

6.5 Data Management

6.5.1 Management and Coordination

As an organization that routinely collects, manages, integrates, and distributes data, EMAP can provide leadership, support, and standardization for related databases.

6.5.2 Tasks

Efforts will include improvement of EMAP data management capabilities, and assistance with management of ecological databases that have no long-term stewards. EMAP has identified several databases useful to EMAP for which long-term stewardship is uncertain, and will assist with technical expertise, and standards.

Modify Inventory of Monitoring Programs Database

AED developed this database for the MAIA Community-Based Assessment Team (CBAT) and will modify it for the Western Pilot.

Design and Develop Aquatic Mortality Database for Atlantic and Gulf Coasts

See Section 2.3.5.1 and Appendix B.8.2

Incorporate Sections of Health, Ecological, and Economic Dimensions of Global Change (HEED) Database

To be determined

Design and Develop the EMAP Archival, Preservation, and Tracking System

The EMAP Archival, Preservation, and Tracking System (EAPTS) will be revised and updated.

Update Oracle Database Design for EMAP-Estuaries Atlantic Coast Data

EMAP Estuaries Resource Group data collected for the Atlantic coast were transferred from SAS data sets to an Oracle database in 1994–95, which is no longer active. The database will be redesigned to improve performance and maintenance, and Atlantic coast data sets (Virginian, Carolinian, MAIA) will be entered into it.

Load EMAP-Estuaries Atlantic Coast Oracle Database With Virginian Province, MAIA, and Carolinian Province Data

EMAP Estuaries Resource Group 1990–1995 data will be converted to an Oracle database (from current SAS format) for long-term management and dissemination.

6.5.3 Responsibilities

For the selected databases, EMAP–IM (AED) will be responsible for design, implementation, maintenance, standards development, and preparation of summary data and metadata.

6.5.4 Resources Needed

To be covered in existing in-house ADP contractor services.

6.6 GIS Spatial Data and Analyses

EMAP–IM (AED) will enhance the functionality of the EMAP–IM system for delivering spatial data and metadata to EMAP users principally by creating a MapObjects capability on the EMAP web sites. Development of these capabilities are outlined below.

6.6.1 Management and Coordination

This task will require coordination with EMAP researchers that generate data (e.g., Landscape Ecology, MAIA), and delivery of data according to FGDC standards. EMAP–IM will also coordinate with CENR to ensure that applicable standards are also followed.

6.6.2 Tasks

Develop MapObjects Capability on EMAP Internal Web Site and Public Access Web Site

6.6.3 Responsibilities

EMAP researchers and GIS staff will generate maps that will be placed on the Web Site. EMAP–IM (AED) will develop the MapObjects functionality for delivering these maps through the web site.

6.6.4 Resources Needed

MapObjects software must be purchased and installed on the AED and RTP servers. AED and ETSD will determine resource allocations.

6.7 Mid–Atlantic Integrated Assessment

A major goal of the MAIA program is to transfer technology and approaches from EMAP information management to the regional studies through pilot implementation of regional information systems compatible with and based on the EMAP Information Management approach. EMAP–IM (AED) is leading development of a prototype IM system for MAIA. The main products of this effort will be a MAIA Information Management Plan, MAIA web site, and MAIA inventory

of monitoring programs already developed by the Community-Based Assessment Team (CBAT). This section describes the tasks that will be implemented to meet these goals.

6.7.1 Management and Coordination

The MAIA web site and system must be compatible with the EMAP-IM system so that the two sites can be cross-referenced and linked. Implementing the components, ensuring compatibility, and transferring relevant system approaches and standards (e.g., database design, metadata standards) to Region III requires EMAP-IM (AED) to coordinate closely with MAIA and Region III's Community-Based Assessment Team.

6.7.2 Tasks

Determine Region III Needs for Data Management

EPA Region III (Philadelphia, PA) has been identified as the location for implementing the information system for MAIA. As part of the pilot project EMAP will participate in establishing information management requirements for the Region and assist with implementing the system based on EMAP-IM principles.

Develop MAIA Information Management Plan

An Information Management Plan will be developed for the MAIA program and Region III.

Develop MAIA Web Site Prototype and Final Design

Prototype web pages have been developed and tested for MAIA on the EMAP Internal Web Site, and contains functionality, navigation and formats in common with the EMAP World Wide Web Site. The MAIA web pages have been transferred to the EPA public access server and linked to the EMAP Public Access Web Site (MAIA 1998c). Development of these web pages is part of the effort to transfer web site technology and approaches to Region III to support their needs in the MAIA program.

MAIA Inventory of Monitoring Programs Online

EMAP will make the MAIA Inventory of environmental monitoring and research programs (MAIA 1998d) available on the EMAP World Wide Web Site MAIA. The Inventory has already been developed by the MAIA Community-Based Assessment Team (CBAT) under CENR's National Environmental Monitoring Initiative (NEMI) (CENR 1998b). The Inventory is an Oracle database that tracks environmental monitoring programs in the MAIA geographical region. (Information about data sets produced by these monitoring programs could be entered into the EMAP Data Directory and cross-referenced to the Inventory.)

Technology Transfer From MAIA to Western Pilot

EMAP will begin transfer of IM approaches and tools from MAIA to the Western pilot project. The transfer will include the MAIA Information Management Plan and details of the implementation for both EMAP and the Region III information systems.

6.7.3 Responsibilities

AED and CBAT will lead this effort.

6.7.4 Resources Needed

Outside contractor services are in place.

6.8 Western Pilot Regional Assessment

It is expected that the information management technology developed and the lessons learned from the EMAP–IM system and the MAIA IM system will be transferred to the Western pilot (see Section 2.3.1, ORD-Regional-Scale Assessments Program, for more information on the Western pilot).

6.8.1 Management and Coordination

One important goal of EMAP–IM in the ORD-Regional Assessment effort is to transfer information management capabilities developed in EMAP to organizations responsible for regional monitoring and assessment (e.g., EPA regions, state and local government agencies). This technology transfer will be piloted with EPA Region III in the MAIA project. Based on this effort, a plan for technology transfer to the regional programs will be developed and presented to the IMWG for review. Once the plan is approved, it will be used to support the development of regional capabilities for managing monitoring data using an approach compatible with EMAP and with CENR. The upcoming Western pilot regional assessment will benefit from technology transfer from EMAP and MAIA.

6.8.2 Tasks

- Determine Regions VII, VIII, IX, and X needs for data management to conduct regional environmental assessments;
- Develop Western pilot Information Management Plan;
- Develop Western pilot Web site;
- Prototype;
- Final design; and
- Western pilot Inventory of Monitoring Programs online.

Section 6, Implementation Plan

- Revise Western pilot Information Management Plan; and
- Update Western pilot Web site.

6.8.3 Responsibilities

AED will take the lead on coordinating information management.

6.8.4 Resources Needed

Data management support will be provided by ORD NHEERL/NERL Divisions in Narragansett, Corvallis, and Las Vegas. They will need additional in-house ADP contract staff, hardware, and software. Support of data sources will be needed to move toward common polices and standards (funds for data management personnel, training, etc.). This could include augmenting existing data management capabilities in Regions VIII, IX, and X. Contract support will also be needed to help conduct data management needs analysis and to prepare an Information Management Plan.

6.9 EMAP–IM System Administration and Coordination

6.9.1 Tasks

As the lead group for EMAP information management, EMAP–IM (AED) will conduct a number of project administration and coordination activities, as described in the tasks below.

Begin Coordinating Western Pilot Data Management

EMAP–IM will begin introducing information management issues and coordination for the planned Western Pilot. EMAP–IM (AED) will coordinate with relevant program initiators and contribute to project planning efforts, including:

- Form Western pilot data management working group and supply the group with existing EMAP guidelines, procedures, and support necessary to apply current EMAP–IM approach described in this Plan; and
- Conduct session on data management at EMAP Western Ecoregions Symposium to initiate technology transfer and establish existing systems and approaches for information management in this region.

Distribute 1998 EMAP Information Management Plan

The latest EMAP Information Management Plan will be publicly distributed as an EPA publication.

Section 6, Implementation Plan

Revise EMAP Data Policies, Guidelines, and Standards

EMAP-IM (AED) will continue enhancing EMAP information management policies, guidelines and standards. Review and necessary revision of policies, guidelines, and standards will be conducted annually by EMAP-IM in 1999, 2000, and 2001.

Conduct Information Management Session at MAIA Working Conference (Baltimore, MD)

EMAP-IM (AED) will conduct a session on EMAP information management, which will be an opportunity to discuss progress and strategy of regional information management with the researchers conducting the work.

Workshop on Estuarine Data Management at Estuarine Research Federation Meeting (New Orleans, LA)

Workshops to discuss management of estuarine data will be organized for the 1999 and 2001 Estuarine Research Foundation (ERF) meetings.

Produce Next Version of EMAP Information Management Plan

A revised version of this EMAP Information Management Plan will be produced in 2001 that updates all of the activities described herein.

6.10 Overall Resource Requirements for EMAP-IM (AED)

EMAP-IM members (including EMAP-IM (AED), Resource Groups, and Working Groups) must have adequate resources and expertise for meeting the information management goals and standards outlined in this Information Management Plan.

EMAP-IM (AED) requires the following professional support to complete the implementation tasks for the next three years:

- Chair of the IMWG—to oversee the system development;
- Database specialist—to design, implement and maintain the relational or object-oriented databases;
- Internet system specialist—to support the development and implementation of the web site;
- Programmer/system tester—to develop interfaces and to test integrated system;
- Data Librarian —to support development of documentation and to maintain physical holdings of all files. The data librarian will also support the location and access of both data and metadata;

Section 6, Implementation Plan

- Data Analyst—to support data analyses, aggregations and presentation. Includes development of graphic summaries for appropriate presentation; and
- GIS Analyst/Programmer—to develop GIS coverages, conduct spatial analyses, and develop map-based query application on web site.

Sufficient funding for travel needs to be included to support meetings with the wide variety of partners and users involved in this program. AED participation in standard development and technology transfer is critical to the continued success of the EMAP–IM approach.

6.11 Conclusions

The tasks described in this Implementation section support EMAP–IM’s efforts to implement a successful decentralized information management approach (see Section 1, Introduction and Approach). Over time, the list of tasks will be modified as EMAP responsibilities evolve in response to new and expanded efforts.

6.12 Tentative Schedule

6.12.1 FY1999 Tasks

Administration and Coordination

Begin coordinating Western pilot data management	Oct 1998
Form Western pilot data management working group	Oct 1998
Distribute EMAP Information Management Plan	Dec 1998
Revise EMAP data policies, guidelines, and standards	Feb 1999
Session on Information Management at MAIA Working Conference, Baltimore, MD	Dec 1998
Session on data management at EMAP Western Ecoregions Symposium	Apr 1999
Workshop on estuarine data management at Estuarine Research Federation meeting, New Orleans, LA	Sep 1999

EMAP Web Site

Begin redesign to incorporate more database capabilities	Dec 1998
Begin adding MapObjects capability to internal site	Nov 1998
Add capability for partners outside EPA to access preliminary data	Jan 1999

Section 6, Implementation Plan

Update Web publishing policy, guidelines, and standards manual	Feb 1999
Data Directory	
Move Data Directory Oracle database to public EMAP Web site	Dec 1998
Review revised EIMS data directory/catalog	Oct–Nov 1998
Begin transfer of EMAP Data Directory entries to EIMS	Feb 1999
Request that EPA Web site add Z39.50 protocol	Oct 1998
Revision of Oracle database	Apr–May 1999
Revision of Oracle client (Data Directory query, mgmt tool)	Jun 1999
Update Data Directory policy, guidelines, and standards manual	Feb 1999
Data Catalog	
Evaluate migration from WordPerfect template to specialized metadata software	Mar 1999
Update Data Catalog policy, guidelines, and standards manual	Feb 1999
EMAP Bibliographic Database	
Begin cleaning data	Oct 1998
Revise Web-based query form	Oct 1998
Write data submission format guidelines	Nov 1998
Move EMAP Bibliographic Database to public EMAP Web site	Mar 1999
Data Management	
Design and develop the EMAP Archival, Preservation, and Tracking System	Jan 1999
Update Oracle database design for EMAP-Estuaries Atlantic Coast data	Jan–Apr 1999
Load EMAP-Estuaries Atlantic Coast Oracle database with Virginian Province, MAIA, and Carolinian Province data	Apr–Jun 1999
Mid-Atlantic Integrated Assessment (MAIA)	
Determine Region III needs for data management	Oct 1998
Develop MAIA Information Management Plan	Nov 1999
Develop MAIA Program Web site	

Section 6, Implementation Plan

Prototype	Oct 1998
Final Design	Dec 1999
MAIA Inventory of Monitoring Programs online	Dec 1998
Tech transfer from MAIA to Western pilot	Apr 1999
EMAP 1990–1995 Data and Metadata	
Get remaining Resource Groups data	
Forests	As rec'd
Agro-ecosystems	As rec'd
Get remaining Round 1 R-EMAP data	As rec'd
EMAP 1996–2001 Data and Metadata	
MAIA—Estuaries	
Finish loading 1997 data	Feb 1999
Load 1998 data	Dec 1999
MAIA—Surface Waters	
Finish loading 1997 data	Feb 1999
Load 1998 data	Dec 1999
MAIA—Landscape Ecology	
Load landscape indicator coverages	Dec 1998
UVB data (NERL)	
Finish database design	Oct 1998
Finish Web site design	Oct 1998
R-EMAP	
Load data from Round 2 studies	As rec'd
6.12.2 FY2000 Tasks	
Administration and Coordination	
Revise EMAP data policies, guidelines, and standards	Feb 2000
EMAP Web Site	
Develop MapObjects capability on EMAP public Web site	Nov 1999
Review by EMAP IMWG	Jan 2000
Update Web publishing policy, guidelines, and standards manual	Feb 2000

Section 6, Implementation Plan

Data Directory

Synchronize with U.S. Global Change Research Program data directories	Nov 1999
Revise directory to be compatible with Z39.50 protocol	Oct–Nov 1999
Load Western pilot directory entries	Jan 2000
Update Data Directory policy, guidelines, and standards manual	Feb 2000

Data Catalog

Update Data Catalog policy, guidelines, and standards manual	Feb 2000
--	----------

EMAP Bibliographic Database

Load Western pilot bibliography	Oct 1999
Update data submission format guidelines	Apr 2000

Data Management

Revise the EMAP Archival, Preservation, and Tracking System	Oct–Dec 1999
---	--------------

GIS Spatial Data and Analyses

Develop MapObjects capability on EMAP World Wide Web Site	Oct–Dec 1999
---	--------------

Western Pilot

Determine Regions VII, VIII, IX, and X needs for data management	Feb–Apr 1999
Develop Western pilot Information Management Plan	Jun 1999
Develop Western pilot Web site	Dec 1999
Western pilot Inventory of Monitoring Programs online	Dec 1999

EMAP 1996–2001 Data and Metadata

MAIA-Estuaries	
Finish loading 1998 data	Dec 1999
MAIA-Surface Waters	
Finish loading 1998 data	Dec 1999
UVB data (NERL)	
Load FY2000 data	Jun–Aug 2000
R-EMAP	
Load data from FY1999 studies	Jun–Aug 2000

Section 6, Implementation Plan

6.12.3 FY2001 Tasks

Administration and Coordination

Revise EMAP data policies, guidelines, and standards	Feb 2001
Workshop on estuarine data management at Estuarine Research Federation meeting	Sep 2001
Produce next version of EMAP Information Management Plan	Sep 2001

EMAP Web Site

Update Web publishing policy, guidelines, and standards manual	Feb 2001
--	----------

Data Directory

Revision of Oracle database	Apr–May 2001
Revision of Oracle client (Data Directory query, mgmt tool)	Jun 2001
Update Data Directory policy, guidelines, and standards manual	Feb 2001

Data Catalog

Update Data Catalog policy, guidelines, and standards manual	Feb 2001
--	----------

EMAP Bibliographic Database

Load new entries	Oct 2000
------------------	----------

Data Management

Update the EMAP Archival, Preservation, and Tracking System	Oct–Dec 2000
Update Oracle database designs	Jan–Mar 2001

Western pilot

Revise Western pilot Information Management Plan	Nov 2000
Update Western pilot Web site	Dec 2000

EMAP 1996–2001 Data and Metadata

Western pilot	
Load remaining data	Jun–Aug 2001
UVB data (NERL)	
Load FY2001 data	Jun–Aug 2001
R-EMAP	
Load FY2000 data	Jun–Aug 2001

References

- AIRMon. 1998 (June). Atmospheric Integrated Research Monitoring Network.
<http://nadp.sws.uiuc.edu/AIRMoN/>
- AIRS. 1998 (June). AIRS database. <http://www.epa.gov/airsweb/>
- Barton, G. 1996. NOAA Environmental Services Data Directory. Earth Systems Monitor, December 1996. 6–8.
- Barton, G. 1997. NOAA and the Federal Geographic Data Committee: Earth Systems Monitor 7(3), March 1997.
- Buffum, H. and S. Hale. 1997a. MAIA-Estuaries 1997 laboratory data format and transmittal guidelines. Atlantic Ecology Division, NHEERL, U. S. Environmental Protection Agency. Narragansett, RI. 11 p.
- Buffum, H. and S. Hale. 1997b. MAIA-Estuaries 1997 summary database data format and transfer guidelines. Atlantic Ecology Division, NHEERL, U. S. Environmental Protection Agency. Narragansett, RI. 33 p.
- CASTNet. 1998 (June). CASTNet database.
<http://www.epa.gov/ardpublic/acidrain/castnet/about.html>
- CBP. 1998 (July). Chesapeake Bay Program web site.
<http://www.chesapeakebay.net/bayprogram/>
- CENR. 1994. “The U.S. Global Change Data and Information System Implementation Plan.” Report, Committee on Environment and Natural Resources, National Science and Technology Council, Washington, D.C.
- CENR. 1998a (June). CENR information on White House Home Page.
<http://www.whitehouse.gov/WH/EOP/OSTP/NSTC/html/enr/enr-plan.html>
- CENR. 1998b (June). CENR National Environmental Monitoring Initiative overview.
<http://www.epa.gov/monitor/>

References

- CENR. 1998c (June). CENR overview at NOAA. <http://www.nnic.noaa.gov/CENR/cenr.html>
- Chinn, H. and Bledsoe, C. 1997. "Internet access to ecological information—the US LTER All-Site Bibliography Project," *BioScience* 47(1), :50–57.
- CIESIN. 1998 (June). CIESIN gateway to EPA databases.
<http://epawww.ciesin.org/national/epahome/epahome.html>
- CISNet. 1998 (June). CISNet program information. <http://es.epa.gov/ncerqa/rfa/ncoast.html>
- DISPro. 1998a (June). DISPro overview. <http://www.epa.gov/emap/html/dispro.html>
- DISPro. 1998b (June). DISPro gaseous pollutants .
<http://www.aqd.nps.gov/ard1/gas/dispro1.htm>
- EEI. 1998 (June). EPA OIRM Essential Elements of Information Requirements.
<http://www.epa.gov/irmpoli8/sysdesn/>
- EIMS. 1998. (June). EPA Environmental Information Management System.
<http://www.epa.gov/eims/>
- EMAP. 1998. (June). EMAP World Wide Web Site. <http://www.epa.gov/emap/>
- ENVIROFACTS. 1998 (July). EPA Envirofacts databases. <http://www.epa.gov/enviro/>
- ERIN. 1998 (June). ERIN, Australian national monitoring data web site. <http://www.erin.gov.au/>
- ESA. 1995. Report of the Ecological Society of America Committee on the Future of Long-Term Ecological Data. Vol. 1. <http://www.sdsc.edu/~ESA/>
- FGDC. 1994. Content Standards for Digital Geospatial Metadata, June 8, 1994, Federal Geographic Data Committee, Washington, DC.
- Frithsen, J. B. 1996a. "Suggested modifications to the EMAP data set directory and catalog for implementation in US EPA Region 10. Draft, June 11, 1996." Report prepared for the U.S. Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC., by Versar, Inc., Columbia, MD.
- Frithsen, J. B. 1996b. "Directory Keywords: Restricted vs. unrestricted vocabulary. Draft, May 21, 1996," Report prepared for the U.S. Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC., by Versar, Inc., Columbia, MD.

References

- Frithsen, J. B., and D. E. Strebel. 1995. "Summary documentation for EMAP data: Guidelines for the Information Management Directory. 30 April 1995," Report prepared for U.S. Environmental Protection Agency, Environmental Monitoring and Assessment Program (EMAP), Washington, DC. Prepared by Versar, Inc., Columbia, MD.
<http://www.epa.gov/emap/html/infomgmt.html>
- Gane, C. and T. Sarson. 1980. *Structured Systems Analysis: Tools and Techniques*. McDonnell Douglas, St. Louis, MO.
- GCRIO. 1998 (July). *Policy Statements on Data Management for Global Change (1991) Research*. <http://www.gcdis.usgcrp.gov/data-policy.html>
- GCRP. 1992. *USGC Data and Information Program Plan*. U.S. Global Change Research Program. Committee on Environment and Natural Resources, National Science and Technology Council, Washington, D.C.
- GCRP. 1995a. *GCDIS Implementation 1995. Vol. I—Interagency Implementation*. U.S. Global Change Research Program. Committee on Environment and Natural Resources, National Science and Technology Council, Washington, D.C.
- GCRP. 1995b. *GCDIS Implementation 1995. Vol. II—Agency Implementation*. U.S. Global Change Research Program. Committee on Environment and Natural Resources, National Science and Technology Council, Washington, D.C.
- GCRP. 1998. "An Investment in Science for the Nation's Future—A Report by the Subcommittee on Global Change Research, Committee on Environment and Natural Resources of the National Science and Technology Council: A Supplement to the President's Fiscal 1998 Budget." Washington, D.C.
- GMNET. 1998 (July). *Gulf of Mexico Program's Gulf of Mexico Aquatic Mortality Network web site*. <http://pelican.gmpo.gov>
- GISTools. 1998 (June). *EPA GIS Tools home page*. <http://www.epa.gov/epahome/gis.htm>
- Hale, S.S., M.H. Hughes, J.F. Paul, R.S. Mcaskill, S.A. Rego, D.R. Bender, N.J. Dodge, T.L. Richter, and J.L. Copeland. 1998. *Managing scientific data: The EMAP approach*. *Environmental Monitoring and Assessment* 51:429-440.
- IMPROVE. 1998. *Interagency Monitoring of Protected Visual Environments data site*. <ftp://vista.cira.colostate.edu>.

References

- ITIS. 1998 (July). Integrated Taxonomic Information System (ITIS) web page at USGS.
<http://biology.usgs.gov/cbi2/programs/itlis.html>
- LTER. 1995. Draft proceedings of the 1995 Long-Term Ecological Research Data Management Workshop, July 27-29, 1995, Snowbird, Colorado.
- MAIA. 1998 (June). MAIA Home Page. http://www.epa.gov/docs/grd/maia_overview.html
- MDN. 1998 (June). Mercury Deposition Network: a NADP/NTN Subnetwork.
<http://nadp.sws.uiuc.edu/mdn/>
- MRLC. 1998a (June). Multi-Resolution Land Characteristics Consortium Home Page.
<http://www.epa.gov/docs/grd/mrlc/>
- MRLC. 1998b (June). Multi-Resolution Land Characteristics Consortium overview.
<http://www.epa.gov/mrlc/Databases.html>
- NADP/NTN. 1998 (June). National Atmospheric Deposition Home Page.
<http://nadp.sws.uiuc.edu/default.html>
- NASA. 1991. "Directory Interchange Format Manual; Version 4.0," NASA, National Space Science Data Center, Greenbelt, MD. December 1991.
- National Academy of Sciences. 1990. Managing Troubled Waters. Committee on a Systems Assessment of Marine Environmental Monitoring. Washington, D.C.
- NBII. 1998. National Biological Information Infrastructure MetaMaker web site.
http://www.emtc.nbs.gov/http_data/emtc_spatial/applications/nbiimker.html
- NDDN. 1998. National Park Service Dry Deposition Network web site. <http://www.aqd.nps.gov/ard1/gas/nddn1.htm>
- NOAA. 1996. Metadata resources fact sheet. NOAA Coastal Services Center, Charleston, SC.
- NPS. 1998a (June). NPS Air Monitoring Programs web site.
<http://www.aqd.nps.gov/ard1/proghp.html>
- NPS. 1998b (June). NPS Air Resource Division. <http://www.aqd.nps.gov/ard/>
- NPS. 1998c (June). NPS Dry Deposition Monitoring Network.
<http://www.aqd.nps.gov/ard1/gas/nddn1.htm>

References

- NPS. 1998d (June). NPS Inventory and Monitoring Program. <http://www.aqd.nps.gov/nrid/im/>
- NRC. 1994. Review of EPA's Environmental Monitoring and Assessment Program: Forest and Estuaries Components. Pre-Publication Draft. National Research Council, Washington DC.
- NRC. 1995a. "Finding the forest in the trees: The challenge of combining diverse environmental data," National Academy Press, Washington, DC. 129 pp.
- NRC. 1995b. Promoting the National Data Infrastructure through Partnerships. National Research Council.
- NRCS. 1998 (June). U.S. Natural Resources Conservation Service Natural Resources Inventory. <http://www.nhq.nrcs.usda.gov/NRI/intro.html>
- NUVMC. 1998 (June). National UV Monitoring Center (University of Georgia). <http://oz.physast.uga.edu/>
- REMAP. 1998a. REMAP Projects 1990-1995. <http://www.epa.gov/emap/html/remap1.html>
- REMAP. 1998b. REMAP Projects 1996. <http://www.epa.gov/emap/html/remap2.html>
- ReVA. 1998a (June). ReVA-Landscape Ecology San Pedro River Project Description. <http://www.epa.gov/crdlvweb/land-sci/san-pedro.htm>
- ReVA. 1998b (June). ReVA-Landscape Ecology Semi-Arid Land-Surface-Atmosphere (SALSA) Program Cooperative Research in Global Change Processes in Semi-Arid Lands. <http://www.tucson.ars.ag.gov/salsa/salsahome.html>
- ReVA. 1998c. ReVA home page. <http://www.epa.gov/nerlrva/index.html>
- Riitters, K. H. & J. D. Wickham, 1995. A Landscape Atlas of the Chesapeake Bay Watershed, Tennessee Valley Authority and U.S. EPA, December 1995.
- SCCWRP. 1998. Pilot Project Survey. <http://www.sccwrp.org/data/datalink.htm>
- Shepanek, R. 1994. EMAP Information Management Strategic Plan: 1993B1997. EPA/620/R-94. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC
- STORET. 1998 (June). STORET database. <http://epaserver.ciesin.org/national/epaorg/storet.html>

References

- Strebel, D. E., and J. B. Frithsen. 1995a. "Guidelines for distributing EMAP data and information via the Internet. April 30, 1995," Prepared for U.S. Environmental Protection Agency, Environmental Monitoring and Assessment Program, Washington, DC. Prepared by Versar, Inc., Columbia, MD. <http://www.epa.gov/emap/html/infomgmt.html>
- Strebel, D. E., and J. B. Frithsen, 1995b. "Scientific documentation for EMAP data: Guidelines for the information management catalog. Draft: April 30, 1995," Prepared for U.S. Environmental Protection Agency, Office of Modeling, Monitoring Systems and Quality Assurance, Washington, DC. Prepared by Versar, Inc., Columbia, MD. <http://www.epa.gov/emap/html/infomgmt.html>
- SURF. 1998 (June). Surf Your Watershed web site. <http://www.epa.gov/surf/>
- UK Environmental Change Network. 1998 (June). United Kingdom Environmental Change Network web site. <http://mwnta.nmw.ac.uk/eddemo/>
- U.S. EPA. 1991. EEI-1 Mission Needs Statement EMAP. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.
- U.S. EPA. 1993a. EMAP Information Management Task Group, System Life Cycle Management Studies Manual, (draft), U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U. S. EPA. 1993b. EMAP IM POC Standards Manual (draft), U.S. Environmental Protection Agency.
- U. S. EPA. 1993c. Environmental Monitoring and Assessment Program: Master Glossary. EPA/620/R-93/013, U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.
- U.S. EPA. 1993d. System Design and Development Guidance. EPA Directive Number 2182. United States Environmental Protection Agency.
- U.S. EPA. 1994a. Data Catalog and Dictionary. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U.S. EPA. 1994b. User's Guide for EMAP IM System—12/8/94. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

References

- U.S. EPA. 1994c. Landscape Monitoring and Assessment Research Plan. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/620/R-94/009
- U.S. EPA. 1995a. Strategic Plan for the Office of Research and Development, November 1995 External Review Draft. EPA/600/R-95/162. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U.S. EPA. 1995b. "Providing information to decision makers to protect human health and the environment," Information Resources Management Strategic Plan. EPA-220-B-95-002. April 1995. U.S. Environmental Protection Agency, Administration and Resources Management, Washington, DC.
- U.S. EPA. 1995c. Information Resources Management Policy Manual. United States Environmental Protection Agency, Administration and Resources Management, Washington, DC.
- U.S. EPA. 1995d Mid-Atlantic Landscape Indicators Project Plan U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/620/R-95/003
- U.S. EPA. 1996a. EMAP Information Management Plan, October, 1996 draft. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U.S. EPA. 1996b. "EMAP Research Plan. Draft, July 1996," U.S. Environmental Protection Agency, Office of Research and Development, NHEERL, Research Triangle Park, NC.
- U.S. EPA. 1996c. "ORD Information Management Strategic Plan," U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U.S. EPA. 1996d. "Providing information to decision makers to protect human health and the environment. Information Resources Management Five-Year IRM Implementation Plan, February 1996 draft," U.S. Environmental Protection Agency, Administration and Resources Management, Washington, DC.
- U.S. EPA. 1996e. EMAP Data Management Review Team Report, July 2, 1996 draft. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
- U.S. EPA. 1996f. Addendum to: "Guidelines for the information management directory," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.

References

- U.S. EPA. 1996g. Addendum to: "Guidelines for the information management catalog," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.
- U.S. EPA. 1996h (July). Directive 2100—Information Resource Management [IRM], Chapter 10—Records Management.
- U.S. EPA. 1997a. "EMAP Research Strategy," U.S. Environmental Protection Agency, Office of Research and Development, NHEERL, Research Triangle Park, NC.
- U.S. EPA. 1997b. "EMAP Research Plan," U.S. Environmental Protection Agency, Office of Research and Development, NHEERL, Research Triangle Park, NC.
- U.S. EPA. 1997c. Update to: "ORD's strategic plan." EPA/600/R-97/015. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.
- U.S. EPA, 1998a (in prep.). Ecological Indicator Evaluation Guidelines, U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.
- U.S. EPA. 1998b. Update to: "Guidelines for distributing EMAP data and information via the Internet," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.
- U.S. Forest Service. 1998 (in prep.). <http://www.fs.fed.gov>.
- U.S. Geological Survey. 1998 (June). Formal metadata C Information and tools available on this server. <http://geochange.er.usgs.gov/pub/tools/metadata>
- Williams, N. 1997. "How to get databases talking the same language," *Science* 275: 301-302.
- WOUDC. 1998 (June). World Ozone and UV Radiation Data Centre. <http://www.tor.ec.gc.ca/woudc/woudc.htm>
- Z39.50. 1998 (June). Library of Congress site for Z39.50. <http://lcweb.loc.gov/z3950/agency/>

Glossary

Definitions adapted from U.S. EPA (1993c) and Gane and Sarson (1980)

accuracy	The degree to which a calculation, a measurement, or set of measurements agree with a true value or an accepted reference value.
aggregation	The process of deriving new or <i>summary data</i> by integrating or combining other data sets.
aggregate data	A data set derived from aggregation.
agroecosystem	A dynamic association of crops, pastures, livestock, other flora and fauna, atmosphere, soils and water.
ancillary data	Data collected from studies within EMAP but not used directly in the computation of an <i>indicator</i> .
anonymous FTP	An interactive service provided by many Internet hosts allowing any user to transfer documents, files, programs, and other archived data using FTP.
Arc/Info	Geographic information systems software of the Environmental Systems Research Institute (ESRI).
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 -40EC to 50EC. The vegetation is dominated by woody perennial, succulents, and drought resistant trees.
ASCII	The American Standard Code for Information Interchange set of character codes that allow data to be stored as plain text in a format readable by most computer programs.
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

Glossary

	a socially defined value, and association among indicators of ecological condition and stressors.
attribute	A <i>data element</i> that holds information about an <i>entity</i> .
batch-loaded	A procedure by which a number of data files are loaded into a database by a program containing commands without a need for user interaction during the procedure.
browser	See <i>web browser</i> .
CD-ROM	Compact Disk-Read Only Memory (read-only refers to the fact that material on the disk cannot be modified, but only read by the user).
Census TIGER	Spatial information format used for Census Bureau data.
Central EMAP-IM	The Central EMAP-IM Coordinating Group in the early EMAP program (1990-1995) that led information management and conducted basic information management research in EMAP. The group was composed of scientific and administrative personnel headed by a technical coordinator.
characterization	Determination of the attributes of resource units, populations, or sampling units.
clearinghouse	A web site containing cross-referenced (indexed) information about data important to the user (e.g., an FGDC clearinghouse containing metadata for and <i>hyperlinks</i> to spatial data available at a site on the Internet).
client/server	A system composed of distributed computer systems that are clients to a central server that holds common software tools and data.
completeness	The amount of valid data obtained compared to the planned amount.
conceptual	Abstract or generalized.
condition	In EMAP, the distribution of scores describing resource attributes without respect to any societal value or desired use, that is, a state of being.
condition indicator	A characteristic of the environment that provides quantitative estimates of the state of ecological resources and is conceptually tied to a value.

Glossary

Current EMAP	The re-directed EMAP program (1995–) that is currently being conducted through the Working Groups.
database	An <i>entity</i> or set of entities that hold data (e.g., spreadsheet, relational database, object-oriented database).
data catalog	In EMAP, files containing detailed <i>metadata</i> about data sets.
data dictionary	A description of the format and content of fields in a database. The dictionary is usually formatted as a list or table.
data directory	In EMAP, a listing of summary documentation about data sets (e.g, limited to subject matter, data location, contact names) that is less comprehensive than metadata.
data element	The smallest unit of data that is meaningful for the purpose at hand (e.g., item or field).
data repository	Any computer, web site, or manual system in which data are stored for long-term access and archiving.
data set	A grouping or collection of similar or related data.
data source	In EMAP, an organization or individual that collects or creates data.
distributed information system	Physically separated set of computer or manual systems for managing data and information.
documentation	In EMAP, metadata and Data Directory information that describe data sets.
download	The process of transferring files across a network from an information repository to the user (e.g., from a web site to a personal computer).
Early EMAP	The original EMAP program conducted from 1990–1995, its data, and the Resource Groups.
ecological indicator	See <i>condition indicator</i> .
ecoregion	Regions of relative homogeneity in ecological systems or in relationships between organisms and their environments.
ecosystem	The interacting system of a biological community and its non-living environmental surroundings.

Glossary

EMAP IM Coordinating Group	See <i>Central EMAP-IM</i> .
EMAP-IM (AED)	EMAP Information Management staff at the ORD Atlantic Ecology Division that leads management of data and information for the program.
EMAP-IM	All of the groups involved in EMAP Information Management, including EMAP-IM (AED), Working Groups, and the EMAP IMWG.
EMAP Internal Web Site	The EMAP web pages on the Atlantic Ecology Division internal EPA web server in Narragansett, Rhode Island. Intended for testing and development of EMAP web pages, databases, information products. Access restricted to EPA users.
EMAP Public Access Web Site	The EMAP web pages on the EPA RTP public access web server. Intended for data and information that have passed through the EMAP WWW guidelines (Strebel, D. E., and J. B. Frithsen. 1995a). Accessible by all potential users inside and outside EPA.
entity	A <i>data object</i> in a database.
entity relationship diagram	Graphical representation of <i>data objects</i> and their relationships.
environment	The sum of all external conditions affecting the life, development, and survival of an organism.
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.
estuary	Region of interaction between rivers and ocean waters, where tidal action and river flow mix fresh and salt water.
extensible system	An automated or manual information system that can be extended without being re-implemented.
extranet	An extranet is a private network (e.g., intranet) that uses Internet protocols and the public telecommunication system to securely share part of an organization's information or operations with outside users.

Glossary

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.
FTP	File Transfer Protocol, a client-server protocol which allows a user on one computer to transfer files to and from another computer over a TCP/IP network.
geographic information system	A collection of computer hardware, software, and geographic data designed to capture, store, update, manipulate, analyze, and display geographically referenced data.
Great Lakes	In EMAP, the resource that encompasses the five Great Lakes, wetlands contiguous to the lakes, and the connecting channels.
GRID	ARC/INFO module for analyzing raster data (e.g., remote sensing images).
heterogeneous	Consisting of dissimilar or diverse constituents. In EMAP, existence of many different types of computers, operating systems, software, and network connections connecting to the same information resources on the Web Site).
hyperlink	A reference (e.g., coded text) from a point in one document to a point in another document (or in the same document). A hyperlink is a cross-reference that displays the target point when the user activates it (e.g. by clicking on it with a mouse). <i>Hyperlinks</i> are usually displayed in some distinguishing way, such as in a color, font or style different from surrounding text.
hypertext	Text containing <i>hyperlinks</i> .
index/indices	A model that integrates data through certain measures (e.g., Index of Biological Integrity).
index of information	In EMAP, a directory containing cross-referenced information that facilitates retrieval of data sets by EMAP users.
IMWG	The EMAP Information Management Working Group, composed of members of the Working Groups and EMAP-IM (AED).
indicator	In EMAP, characteristics of the environment, both abiotic and biotic, that can provide quantitative information on ecological resources.

Glossary

indicator development	The process through which an indicator is identified, tested, and implemented.
integrated assessment	An assessment that combines information to provide more than the sum of the individual pieces of information.
integration	The formation, coordination, or blending of units or components into a functioning or unified whole. In EMAP, integration is 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.
interface	A display on a computer screen that gives the user information about how to interact with a database, program, or other online component.
Internet	A world-wide confederation of computer networks that makes possible the exchange of electronic messages, information, and data files.
Internet domain	A set of network addresses (i.e., approved addresses for accessing a site).
intranet	A network that provides Internet-type services but is only accessible to authorized users within an organization. Generally runs on a server on an internal network. In EMAP, it refers to the EPA internal network, which is only accessible to users with EPA IP (internet protocol) addresses.
inventory	A listing of monitoring programs that are producing data sets (e.g., MAIA Inventory of Monitoring Programs).
landscape	The set of traits, patterns, and structure of a specific geographic area, including its biological composition, its physical environment, and its anthropogenic patterns.
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.

Glossary

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.
link	<i>A hyperlink.</i>
logical	The non-physical, underlying nature of system components.
measurement	In EMAP, quantifiable attribute that is tied to an indicator.
metadata	Data about data that describes content, primary elements, quality, background, and other details of data sets. Scientific metadata is information describing scientific data. Spatial metadata describes geographic data and is now governed by the FGDC spatial metadata standard.
model	Mathematical or physical representation of data or a natural system that accounts for some or all of its known properties.
monitoring	In EMAP, the periodic collection of data that is used to determine the condition of ecological resources.
native formats	Electronic file formats used by data sources to manage their data (e.g., Landscape Ecology uses Arc/Info, WED uses SAS).
network domain	The set of computers and computer connections that connect to a single network.
normalized data	A relation (data file) that has no repeating groups of data.
online	Directly connected to a computer so that user input, output, and data access can take place without further human intervention.
open standards	An open standard is a format that is public and widely used, controlled by a neutral public body, is applicable cross-platform, has no proprietary limitations or features, allowing all implementers to play on a level field. Examples include TCP/IP, HTML, FTP.
open systems	An automated system with characteristics that comply with specified, publicly maintained, readily available standards and can be connected to other systems that comply with these standards.
operating system	The low-level software on a computer that runs user applications, schedules tasks, allocates storage, and handles interfaces to peripheral hardware.

Glossary

orphan data sets	Data sets of interest to EMAP that have no long-term stewards. These data sets may be maintained by EMAP IM if they are considered broadly useful to the program.
physical	The implemented components of a system (e.g., databases, hardware).
pilot	Implementation of a subset of planned research scope that will be tested before evolving, at least in part, into a total research or monitoring program.
precision	The degree to which replicate measurements of the same attribute agree or are exact.
prototype	Development of a subset of system functionality for the purpose of evaluating a planned system. The prototype can be used to investigate user requirements.
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.
quality assured data	Data that have been passed through QA and QC (QA/QC) procedures.
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.
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 directly downloaded from a collection instrument that may or may not be quality assured.
research partners	Those Federal agencies participating with EPA in EMAP, including U.S. Department of Agriculture, U.S. Fish and Wildlife Service, U.S. Bureau of Land Management, National Oceanic and Atmospheric Administration, state and local agencies, regional authorities, academic and research institutions, and others.

Glossary

Resource Group	One of eight ecological entities or ecosystem types in EMAP that shares certain basic characteristics. These are: Estuaries, Great Lakes, Lakes and Streams, Wetlands, Forests, Arid Ecosystems, Agro-ecosystems, and Landscape Characterization.
region	Any explicitly defined geographic area. EPA Regions are any of 10 standard Federal regions. In ORD-Regional Assessments, a region is a geographic area that shares basic characteristics that will be studied as a unit.
relation	A two-dimensional array of rows and columns containing a single value and no duplicate rows (e.g., a table in a spreadsheet).
relational database	A database that organizes normalized data tables with key fields in common that allow users to bring together related data from different tables by selecting them based on the common field.
STORET	Information management system developed by the Office of Water for standardizing and monitoring water data.
STORET X	The modernized version of STORET being deployed in 1998.
stressor	Any physical, chemical or biological factor that can induce an adverse response in an ecological system.
surface water	In EMAP, the inland surface waters consisting of all the nation's lakes (other than the Great Lakes), rivers, and streams.
summary data	Quality-assured data resulting from aggregation or integration.
system	In EMAP, any set of computing and organizational components for managing data and information.
system architecture	The overall logical and physical definition of a system.
system design	Specifications for implementation of an system.
system development life cycle	Process used for fielding a completed system.
T1	High speed data communications trunk line that transmits at 1.544 megabits per second.
T3	High speed data communications trunk line that transmits at 44.736 megabits per second.
technology transfer	The process of sharing technology or procedures validated during an pilot or initial test phase to other operations (e.g., transfer of

Glossary

	program and data management procedures from MAIA to the Western pilot).
trends	The changes in the distribution of scores for condition indicators over multiple time periods.
UNIX	A multi-user general-purpose computer operating system.
validation	The process of substantiating specified performance criteria for data or a system.
verification	The process of ensuring correctness in data.
WAIS	A distributed natural-language information retrieval system that allows clients to retrieve documents (text, multimedia) from a server using keywords. Each WAIS search returns a list of documents, ranked according to the frequency of occurrence of the keyword(s). WAIS offers indexed searching for fast retrieval, and a “relevance feedback” mechanism which allows the results of initial searches to influence future searches. It uses the ANSI Z39.50 service.
web browser	Software that displays the contents of hypertext files and provides a means of navigating from one hypertext file to another, across the Internet or on a single computer.
web site	A collection of related files on a server accessible to the World Wide Web that can be read with web browser software.
wetlands	An area of land that is saturated by surface or ground water with vegetation adapted for life under soil conditions (e.g., swamp, bog, marsh).
Working Group	A collection of research partners inside and outside of EPA collaborating on EMAP-funded monitoring projects and led by an ORD employee.
World Wide Web	An client-server information service on the Internet that uses hypertext technology to deliver data and files to web browsers.
X.25	Internationally-used packet switching communication protocol (ISO. X.25) for public data communications networks.
Z39.50	Information Retrieval Service Definition and Protocol Specification for Library Applications that is used for searching library catalogs and databases over the Internet with WAIS. Z39.50 specifies a set of rules and procedures for the behavior of two

Glossary

systems communicating for the purposes of database searching and information retrieval. It is an internationally recognized open standard that enables communication between systems that run on different hardware and use different software. It was developed to overcome the problems associated with multiple database searching such as having to know the unique menus, command language, and search procedures of each system accessed. The U.S. Library of Congress is the official maintenance agency for this standard, which is officially known as ANSI/NISO Z39.50.