United States
Environmental Protection
Agency

Office of Research and
Development
Washington DC 20460

EPA/620/R-99/001b
June 1999

EMAP Information Management Plan: 1998-2001

Appendices

Environmental Monitoring and Assessment Program

EPA/620/R-99/001b
June 1999

EMAP
Information Management Plan: 1998-2001
Appendices

by

Stephen Hale[1], Jeffrey Rosen[2], Dillon Scott[2],
John Paul[1], and Melissa Hughes[3]

[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

Printed on Recycled Paper

Contents

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
A.3.1 EEI-2 Preliminary Design and Options Document 4
A.3.2 EEI-3 Project Management Plan 6

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.4.1 MAIA-Surface Waters 11
B.4.2 MAIA-Estuaries 14
B.4.3 MAIA–Landscape Ecology 20
B.5 Intensive/Index Sites 20
B.5.1 Intensive/Index Sites Program Overview 20
B.5.2 Demonstration Intensive Sites Project 24
B.5.3 Coastal Intensive Sites Network 29
B.6 Landscape Ecology 29
B.7 Regional EMAP (R-EMAP) 46
B.7.1 R-EMAP National Coordinator's Meeting 46
B.7.2 R-EMAP Region I 53
B.7.3 R-EMAP Region VIII 65
B.7.4 R-EMAP Region IX 69
B.8 Ecological Indicator Development 73
B.8.1 Ecological Indicator Development Guidelines and Documentation 73
B.8.2 Aquatic Mortality Monitoring Database 78
B.9 Committee on the Environment and Natural Resources 80
B.9.1 CENR Information 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.2.1 Option Description 119
D.2.2 Option Analysis 121
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

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

Appendix A
Essential Elements of Information Requirements Report

A.1 Purpose
A.2 EEI-1 Mission Needs Analysis
A.3 EEI-2 and EEI-3 Preliminary Design and Options Analysis
A.3.1 EEI-2 Preliminary Design and Options Document
A.3.2 EEI-3 Project Management Plan

A.1 Purpose

EPA Directive 2182 requires that projects that meeting criteria (e.g., national in scope) that are planning an IM system or enhancements to existing systems must address the requirements outlined in EPA's IRM Policy Manual for Essential Elements of Information documentation. The EEI requirements specify a number of system life cycle planning steps and justifies design and implementation of new or enhanced EPA information systems. Three of the EEI requirements applicable to the EMAP–IM system are fulfilled for this version of the EMAP Information Management Plan: EEI-1, Mission Needs Analysis; EEI-2, Preliminary Design and Options Document; and EEI-3, Project Management Plan. Description of the requirements can be found in Volume A and Volume B of the EPA IRM System Design and Development Guidance (U.S. EPA 1993d; EEI 1998).

Fulfilling EEI-1 through EEI-3 requirements was accomplished by revising the October 1996 Draft Information Management Plan and conducting Requirements Analysis workshops with key EMAP Working Groups. The results of the revision and the interviews are incorporated into the body of this EMAP Information Management Plan, as well as five appendices (Table A-1).

To avoid repetition of material already covered in the body of the Plan, appropriate sections are cited for each EEI documentation requirement. Requirements not applicable to EMAP have been omitted, and the reason for the omission is indicated.

Table A-1. Coverage of EEI Requirements in EMAP IM Appendices

EEI Required Component	EMAP IM Plan Section Name and Number	EEI Requirement Fulfilled
EEI-1, Mission Needs Analysis	Data Management Needs and Practices of EMAP Working Groups, Appendix B	Describes the missions and existing problems that are to be addressed by the proposed system development or enhancement. Includes detailed reports of Requirements Analysis interviews with Working Groups
	Introduction and Approach, Section 1	Presents a summary of the need for system enhancement and an Initial System Concept
	Inventory of EMAP Data, Appendix C	A high-level inventory of the major data types generated and used in EMAP, and their approximate volumes and status
EEI-2, Preliminary Design and Options Analysis	Preliminary Design and Options Document, Appendix D	Presents and analyzes system development options to recommend preferred option
EEI-3, Project Management Plan	Project Management and Coordination, Section 5	Specifies how the preferred option can be implemented (project management needs)

A.2 EEI-1 Mission Needs Analysis

The purpose of the Mission Needs Analysis is to document the need for system enhancement, and lay the foundation for the steps and tasks associated with EEI-2 and EEI-3 (Preliminary Design and Options Document and Project Management Plan). The requirements include the following elements:

- overview of the mission, goals, functions, processes, information flows, and problems that underlie the need for system support and justify the need for a new or enhanced information processing solution;
- list of users who define the system needs;
- description of existing system capabilities;
- identification and specification of the required information flow;
- preliminary specification of management requirements for information flow or information processing, and the outputs and benefits of information for the organization's mission and operation; and
- an Initial System Concept for meeting EMAP needs that includes a preliminary depiction of inputs, outputs, and processes (but does not discuss specific hardware and software solutions).

The main product of the Mission Needs Analysis is a Mission Needs Statement. In this Plan, the Mission Needs Statement has been subdivided into three appendices, as shown in Table A-1, and much of the material is covered in the body of the Plan.

EEI Outline Element	Section or Appendix
1. Background	
For agency and organizational mission requiring system support—	
• Mission/function statement(s)	Section 1.1, 1.7
• Organizational chart with key functions/users identified (list of users)	Section 3.2
• Operational environment	Section 4.5
• Current system description, including manual procedures	Sections 4.1, 4.2, 4.3, 4.4
Evolution of defined need—	
• New program or functions	Sections 1 and 2, Appendix B
• Enhancement/modernization of functioning system, or	Sections 3, 4.7, 4.8, 6
• Current performance mode and limitations/problems	Section 4.7, Appendix B
2. Information Flow and Initial System Concept	
Description/documentation of information flow including—	
• Organizational data flow diagrams	Figures 4-1, 4-2. 4-3, 5-2, 5-3
• Key input processes/documents	Section 2, Appendix B
• Primary data integration/database functions and processes	Section 4
• Key output report types and distribution	Section 4, Appendix B
• "Mock-ups" of key output reports and discussion of their benefits to users	OMITTED (too specific for this project because EMAP outputs consist of the use of the Directory and Web Sites).
• Initial System Concept (ideally one page) and related description	Section 1, Introduction and Approach; Section 4.3, System Concept and Overview of Technical Structure

3. Development/Operational Constraints

User commitment, priority, discipline and budgetary limitations—	Sections 3.2, 5; Appendix B
• Policy or organizational constraints	Section 5; Appendix B
• Information security needs based on system sensitivity	Sections 3.5.3.5, 4.5
• Timing of need	Sections 1, 3, 6
• Interface needs	Sections 3, 4
• Shared data/access constraints	Sections 2, 3, 5; Appendix B
• Stability/flexibility of need	Sections 2, 3, 5; Appendix B
• Initiation of project management plan	OMITTED (already included in Section 5, Project Management and Coordination).

A.3 EEI-2 and EEI-3 Preliminary Design and Options Analysis

The Preliminary Design and Options Analysis is intended to produce two documents specified in the EEI guidance (Volume B): EEI-2, Preliminary Design and Options Analysis and EEI-3, Project Management Plan.

A.3.1 EEI-2 Preliminary Design and Options Document

The purpose of the Preliminary Design and Options Document is to translate management and functional requirements delineated in EEI-1 into operational specifications, identify and develop system options for meeting the requirements, analyze the overall feasibility and cost-effectiveness of the options, and choose a preferred option. The document does not describe a complete detailed system.

The required elements for this document are outlined in Volume B, Appendix A of the EEI Requirements, and includes:

- presentation of options for system development that satisfy the Initial System Concept;
- analysis of the benefits and costs, risks and contingencies of these options; and
- selection and justification of the most cost-effective solution.

An indication of how each requirement has been addressed, and why some components are omitted, is presented below. Most of the required topics are included in Section 4, Technical Design; Section 6, Implementation Plan; and Appendix D, Preliminary Design and Options Document.

Because the EMAP–IM system already exists and many of the options for system enhancements are already in the planning stages, discussion of options is included in the body of the Plan under the appropriate system components and activities. The EEI framework for evaluating the adequacy of existing and planned functionality is used to evaluate the current system (see Section 4.7, System Evaluation), and evaluate implementation of an expansion of the system's database capabilities (see Appendix D, Preliminary Design and Options Document). An overview of other future planned developments is presented in Section 6, Implementation (further analysis of these tasks can be added as it becomes available).

EEI Outline Element	Section or Appendix
1. Introduction	
1.1 Background	Sections 1, 3, 4
1.2 Current System Description	Section 4
1.3 Results of Mission Needs Analysis	Sections 2, 5; Appendix B
1.4 Scope and Purpose	Section 1
2. Option Designs	
2.1 System Concept, Management Requirements and Functional Requirements Summary	Sections 1, 3
2.2 Operational Requirements Summary (General system requirements like security, etc.)	Section 3
2.3 Option Descriptions	
2.3.1 Planned enhancements	Section 6; Appendix D
2.3.2 Option for relational database expansion	Appendix D
3. Options Analysis (life cycle benefits and costs, risks and contingencies)	Sections 4, 6; Appendix D
4. Option Recommendation (With Rationale)	Section 6, Appendix D

A.3.2 EEI-3 Project Management Plan

The Project Management Plan is intended to address management issues relating to implementation of the system. The document must address required resources, scheduling, accountability, organizational issues affecting information flow, data stewardship and distribution, the role of EMAP IM, interaction between Working Groups and EMAP IM, required standards, and other issues.

The recommended outline for EEI-3 is presented below with an indication of where each requirement is documented in the IM Plan. Sections not applicable to EMAP IM have been omitted, and the reason for their omission is indicated. Most of the material is covered in Section 5, Project Management and Coordination, and Section 6, Implementation Plan.

EEI Outline Element	Section or Appendix
1. Introduction	
1.1 Background	Sections 1, 3, 5
1.2 Current System Overview	Section 2
2. System Description	Section 4
3. Project Team and Support	Sections 3, 5
3.1 Roles and Responsibilities	Sections 3, 5
3.2 Configuration Management	Sections 3, 5
3.3 Quality Assurance/Control	Sections 3, 4, 5
3.4 Procurement Plan	OMITTED (system procurement issues not applicable in this version)
4. Project Schedule And Task Description	Section 6
5. Project Budget And Funding	OMITTED (budget and funding issues not addressed in this version)
6. Test Plan Requirements/Constraints	Appendix B.10 (description of Requirements Analysis interview process)
7. Project Constraints	Sections 4.7, 4.8, 5
8. Documentation Standards	Section 4
8.1 Policy Events	OMITTED (policy events not addressed in this version)
8.2 Forms and Clearances	OMITTED (internal Agency procedures) not addressed in this version

Appendix B
Data Management Needs and Practices of EMAP Working Groups

B.1 Purpose
B.2 EMAP and Working Group Mission and Goals
B.3 Requirements Analysis Overview
B.4 ORD Regional-Scale Assessments—Mid-Atlantic Integrated Assessment Pilot
- B.4.1 MAIA–Surface Waters
- B.4.2 MAIA–Estuaries
- B.4.3 MAIA–Landscape Ecology

B.5 Intensive/Index Sites
- B.5.1 Intensive/Index Sites Program Overview
- B.5.2 Demonstration Intensive Sites Project
  - B.5.2.1 UV-B Monitoring
  - B.5.2.2 Other DISPro Initiatives
  - B.5.2.3 Related NPS Data
- B.5.3 Coastal Intensive Sites Network

B.6 Landscape Ecology
B.7 Regional EMAP (R-EMAP)
- B.7.1 R-EMAP National Coordinator's Meeting
- B.7.2 R-EMAP Region I
- B.7.3 R-EMAP Region VIII
- B.7.4 R-EMAP Region IX

B.8 Ecological Indicator Development
- B.8.1 Ecological Indicator Development Guidelines and Documentation
- B.8.2 Aquatic Mortality Monitoring Database

B.9 Committee on The Environment and Natural Resources
- B.9.1 CENR Information

B.1 Purpose

This Needs and Requirements Document fulfills part 1 of the EEI-1 Mission Needs Analysis requirements (see Appendix A, Essential Elements of Information Requirements Report) to:

- provide an overview of the mission, goals, functions, processes, information flows, and problems that underlie the need for system support and justify the need for a new or enhanced information processing solution;
- present a list of users who define the system needs;
- describe existing system capabilities; and
- identify and specify the required information flow.

The document presents summaries of the needs and requirements of EMAP Working Groups and end users that were discussed in Requirements Analysis interviews (workshops and conference calls) with EMAP Working Groups, EMAP–IM (AED), and ORD contacts. A sample of the Landscape Ecology Requirements Analysis interview questions is presented in Table B-3 at the end of this chapter. For Working Groups whose missions and needs are not yet fully defined, additional material will be added in future versions of the Plan when more information is available.

B.2 EMAP and Working Group Mission and Goals

The EMAP mission and goals that underlie the needs and requirements of the EMAP–IM system are summarized in Sections 1.4–1.7 of the Information Management Plan. Mission and goals for the Working Groups are described in Section 2, EMAP Data, and Section B.3, Requirements Analysis Overview.

B.3 Requirements Analysis Overview

The following subsections provide full meeting summaries from the Requirements Analysis interviews of the Working Groups. The questions addressed in the interviews are illustrated in Section B.4. Additional material, including information about data collection and management activities, is summarized in Section 2, EMAP Data. A summary of data types, volumes, and status is in Appendix C, Inventory of EMAP Data.

The principal needs and requirements are those of the EMAP Working Groups as data collectors and users. Additional information about these needs is included in this Appendix and in Sections 2.3 (Current EMAP (1996–2001) Data), 3.2.2 (Primary Users), 3.3 Recommended Guidelines for EMAP Data Sources), and 5.4 (Information Management in Working Groups).

The purpose of interviews was to incorporate the needs of the Working Groups and the requirements for EMAP participation in the CENR framework into the system design. Eleven interviews were conducted during Fall 1997–Spring 1998 as shown in Table B-1.

Interviews included a series of questions designed to understand the enhancements derived from review of the EEI-1 requirements and included the following topics:

- **Data sources and collection**—data-generating projects now funded and their timetables, and their sources of data;
- **Data management**—procedures for data quality and completeness, and maintenance;
- **Data products**—data sets, aggregates, and analytical products, resulting publications, interactions with other research entities, users/audience for finished products;
- **Data flow**—the flow of data and information from collection to QA to analysis to aggregation, milestones of data generation and processing, what flow is needed between Working Groups and between EMAP–IM (AED);
- **Hardware, software, infrastructure**—tools (software, methods) for producing data analyses and aggregates, products from data analysis operational environment, status and frequency of upgrades, required and desired functionality;
- **Documentation**—procedures for producing and disseminating metadata on the quality and details of data sets;
- **Data distribution**—methods and formats for distribution of summary data and metadata, primary and secondary data users, limitations/security on data sharing;
- **Data archiving and storage**—procedures and locations for ensuring long-term maintenance of raw data;
- **Project management**—how project structures affect the success of information management and accessibility;
- **Data standards**—usefulness and appropriateness of EMAP–IM system standards for research partners to follow;
- **EMAP–IM (AED) assistance**—standards, resources, expertise, assistance, and guidance desired from EMAP–IM (AED) to assist Working Groups with information management and accessibility; and
- **EMAP–IM system design**—user perceptions of adequacy of existing system features for meeting requirements for locating pertinent data, and distributing and documenting EMAP data.

Table B-1. Requirements Analysis Interviews

Working Group	Participants (EMAP–IM (AED) & ITAS contractor attended all)	Location and Date
ORD Regional Assessments/ MAIA—Surface Waters	ORD-Western Ecology Division	Conference Call March 16, 1998
ORD Regional Assessments/ MAIA—Estuaries	ORD-Atlantic Ecology Division	Conference Call April 2, 1998
ORD Regional Assessments/ MAIA—Landscape Ecology	See Landscape Ecology	Information covered as part of Landscape Ecology interview, below
Intensive/Index Sites (overview)	ORD-Gulf Ecology Division	Conference Call July 24, 1997
Intensive/Index Sites—Demonstration Intensive Sites Program	National Park Service ORD-Western Ecology Division	Conference Call April 13, 1998
Intensive/Index Sites—Coastal Intensive Sites Network		Project in early stages, no interview conducted
Landscape Ecology	NERL Landscape Ecology staff	2-day workshop at EPA Landscape Ecology Branch, Environmental Sciences Division, National Exposure Research Laboratory, Las Vegas, NV September 24–25, 1997
R-EMAP Coordinator's Conference	Regional R-EMAP Coordinators	Meeting at EPA Laboratory, Research Triangle Park, NC November 7, 1997
R-EMAP Region I	Region I Laboratory R-EMAP staff	2-day workshop at EPA Region I Laboratory, Lexington, MA September 16, 1997
R-EMAP Region VIII	Region VIII R-EMAP Coordinator	Conference Call December 19, 1997
R-EMAP Region IX	Region IX R-EMAP Coordinator	Conference Call November 21, 1997
Indicator Development	ORD-Gulf Ecology Division	Conference Call March 31, 1998
CENR	Tom Mace (Chair, CENR Data Management Working Group)	Reviewed literature and memos; personal communication December 17, 1997

The Working Groups comprise a complex network of research projects with overlapping and distinct data needs. The following sections describe the user types or needs unique to each group.

B.4 ORD Regional-Scale Assessments—Mid-Atlantic Integrated Assessment Pilot

The ORD Regional Scale Assessment Pilot program in the Mid-Atlantic (MAIA) has three major data-generating components—Surface Waters, Estuaries, and Landscape Ecology—which are summarized in separate sections below.

B.4.1 MAIA-Surface Waters

Summary of Conference Call
March 16, 1998

Participants

ORD-Corvallis
- John Stoddard
- Tony Olsen

OAO Corp.
- Marlys Cappaert

ORD-Narragansett
- Stephen Hale

Technology Planning and Management Corporation, ITAS Contractor
- Jeff Rosen
- Dillon Scott

Mission and Goals
To estimate the ecological condition of mid-Atlantic streams.

Data Collection, QA/QC, Analysis, Aggregation
The purpose of data collection in this group is to create "metrics." Metrics are characterizations of watershed condition based on measurement of stream characteristics (physical, chemical, biological). Metrics 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 measures of the distribution of substrate types along a particular stream reach). Metrics are created by combining raw field and laboratory 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)). Raw data include 1000's of parameters collected at 150–200 stream stations every summer from 1993–1997. The data are generated by EPA and non-EPA researchers, who verify and validate them and transmit them to ORD. ORD sends the data to "Indicator Leads," the researchers in charge of developing metrics.

Data continue to be collected for 1998 and into the foreseeable future.

Data management is conducted by a combination of field and laboratory staff, Indicator Leads, and ORD-Corvallis.

Data are stored in SAS. Codes for species and other parameters are unique EMAP codes designed to fit the SAS field limit of eight characters (for example, the first four letters of a genus and first four letters of species are combined to make the code for a particular specimen). Codes are

documented in the metadata. No CAS or NODC codes are used because the researchers need intuitive naming for ease of field data recording. Coding is the responsibility of the Indicator Lead, and EMAP has no procedures in place for updating database codes when species or genus names are changed by taxonomists.

The program needs to address version control, since Indicator Leads and ORD can independently update the same version of a data sets but they do not necessarily track the changes and the resulting differences. They have thought about addressing this problem by setting criteria for when a data set is final.

Data Distribution
Data that will be distributed on the EMAP Public Web Site consists of watershed metrics, which will make up a watershed characteristics database that matches each stream ID with the metrics measured along that stretch of stream. Some of the data will be in SAS files, some in GIS (Arc/Info export) files. Data can also be distributed in spreadsheets on request.

Data Documentation
Metadata are produced in EMAP Data Catalog format. Not all data sets have yet been fully documented.

Descriptions of field and lab data collection and analysis methodology will be placed on the EMAP Public Web Site in Adobe Acrobat format (.PDF). Descriptions of metrics calculations may also be created and placed on the EMAP Public Web Site.

Metadata is targeted to users who have a good technical understanding of the data content (i.e., know the meaning of parameters and data types).

The Working Group has found its metadata sufficient to answer users' questions about data, since they have not received calls asking for clarification (i.e., they have distributed data to other groups such as NAWQA, and there have been no problems).

Data Sources
MAIA Surface Waters uses a number of non-EMAP data sources in their analyses. One data set critical to their efforts is the EPA River Reach (RF3), whose quality varies widely by region. The Region III data had been updated, but corrected data are not available for the whole MAIA area (or did they use RF3 data for the whole MAIA region). Surface Waters also uses data from the Natural Resources Conservation Service's Natural Resources Inventory (NRCS 1998) and Landscape Characterization's MRLC interpreted classified coverages of land cover (MRLC 1998b).

Data Volumes
Volume of data to be made available is anticipated to be approximately 20 MB.

Users
Primary users of MAIA Surface Waters data are Indicator Leads and ORD-Corvallis who analyze and aggregate the data for metrics development. Landscape Ecology and some R-EMAP projects are also using the data in their assessments.

Secondary users include occasional requests from state regulatory agencies, such as:

- New York State DEC requested mercury concentrations in fish from EMAP lake sampling data; and
- The State of Oregon requested temperature data for listing stream according to Clean Water Act 303(d).

The metrics data from the EMAP Public Web Site is adequate for the needs of 95% of anticipated users. There are few requests for raw data, although occasionally a landowner will ask for a list of fish species found along a particular stretch of stream.

Software, Hardware, Infrastructure
These data are maintained by ORD-Corvallis in SAS and Arc/Info on a UNIX server (which will be updated to NT in the future). Tabular data sets are stored in SAS in flat files and GIS data are in Arc/Info.

Project Management
Surface Waters has had little difficulty collecting data from the field/lab staff and Indicator Leads. The only difficulty they have is in getting data sets from other Federal programs. They suggested that part of the difficulty is that production and delivery of data sets and metadata are not considered a recognized deliverable in professional achievements or project management. They contrasted the lack of standards and milestones for data and metadata delivery with the very comprehensive EPA QAPP procedures. QAPP procedures are very comprehensive but only cover data sets until they are through the laboratory process; in order to produce data and documentation for distribution, there are further steps not currently acknowledged in established EPA procedures. It might be helpful to add steps for database delivery and documentation to EPA deliverable requirements, set standards for them, and recognize them as professional achievements.

How can EMAP–IM and the system help them?
They do not find the Data Directory very useful because when they are looking for data sets from other Working Groups, they usually go to that part of the EMAP Public Web Site.

They do not feel that they need assistance with metadata because it is necessary to be very familiar with data content in order to produce it, and parts of it can be produced by running a program they have created that gets min/max values, lat/longs, etc.

Conclusions
MAIA Surface Waters is collecting stream data containing numerous parameters that are being aggregated into metrics that indicate the ecological condition of streams in selected watersheds.

Data collection, verification, and validation are conducted by field and laboratory researchers. Data are delivered to ORD-Corvallis, where metrics and metadata are produced and prepared for distribution on the EMAP Public Web Site. Additional data analysis and aggregation are conducted by Indicator Leads who develop metrics for assessing stream condition.

The Surface Waters group indicated the following obstacles to the success of their work:

- Absence of professional incentives (e.g., professional recognition/penalties) for EPA data sets to be completed and made accessible to users beyond initial researchers;
- Early EMAP data that are not available or do not have sufficient detail (NRI); and
- Inconsistent quality of RF3 data from region to region.

B.4.2 MAIA-Estuaries

Summary of Conference Call
April 2, 1998

Participants

ORD-Atlantic Ecology Division
- Stephen Hale
- John Paul

OAO Corp.
- Harry Buffum

Technology Planning and Management Corporation, ITAS Contractor
- Jeff Rosen
- Dillon Scott

Mission and Goals
The mission of this group is to estimate the ecological condition of mid-Atlantic estuaries. The group is led by Kevin Summers (GED) and John Paul (AED) and consists of a consortium of coastal monitoring organizations (including the Chesapeake Bay Program, NOAA, and others) which have been involved in field monitoring of mid-Atlantic estuaries over extended periods of time.

Data Collection

Data collection was conducted in 1997 and 1998. It focused on filling critical gaps in the early EMAP sampling design and includes the same parameters as the 1990–1995 EMAP–Estuaries data collection, with the addition of nutrients and some toxics (including water quality parameters, benthic infauna, sediment toxicity (1997 only), and fish trawls (1998 only)). Data from the two years is being combined to perform multi-year assessments. Data are collected by a combination of field and research teams from AED and the coastal monitoring consortium. Automated field data systems provided to researchers by AED included data entry screens in Oracle Forms that conduct visual and range checks, and modern tracking systems that include the use of bar-coded sample IDs. However, these tools are only used by a few of the researchers (e.g., ORD-Atlantic Ecology Division, ORD-Gulf Ecology Division, and the National Park Service). Other groups used their own systems, which are mostly manual and are considered by the cooperators to be adequate for their needs.

Data Management and Formats

Data collectors, including AED and the collaborating groups, are responsible for QA/QC, management, and analysis of data they collect. Collaborating groups submit aggregate (summary) data sets to AED in many different file formats. In order to conduct integrated assessments using all of the data, AED will need to re-format incoming data and bring them into an internal MAIA–Estuaries SAS database. Collaborators will submit summary data sets to AED for use in the assessments.

Some of the stations sampled were located outside of the EMAP probability-based sampling design. AED will add the extra stations into the assessments and develop appropriate inclusion probabilities as appropriate. EMAP is sponsoring statistical research to determine options for including probability-based and fixed stations data in the same analyses.

Data Standards

Samples are being analyzed by a variety of laboratories, and data should be submitted in fixed predetermined formats. AED has established data submission, content, and format standards for databases. These standards were approved by the researchers and analytical laboratories in the early stages of the project, but they have not been consistently used. Collaborators submit in many formats, and continue to develop databases and manage data according to their own needs.

Version Control

Version control at AED is now being done by keeping one Read-Only directory on the network server for final versions of data sets. These files are only available to EPA Intranet users. Documentation of the versions and their pedigrees are kept on paper in a file cabinet.

Data Stewardship and Long-Term Maintenance

AED will retain long-term stewardship for the raw data it collects, the data aggregates it generates from data sets of all collaborators, and the documentation for all of these data. Data collaborators

will retain ownership of the raw and processed data they created and will be responsible for its long-term maintenance and accessibility.

Data Aggregation
The AED assessment team will review data submitted by collaborators and creating data aggregates for the assessments. Data aggregates from these assessments will ultimately be posted on the EMAP Public Web Site.

GIS coverages will also be produced. Station location data from the monitoring studies will be entered into GIS coverages. In addition, watersheds will be delineated for the estuaries sampled (by using the USGS Digital Raster Graphs (DRGs)). Overlays of these data will be combined in the assessments with commonly available coverages from the MAIA Geographic Reference Database, MAIA regional collaborators, Landscape Ecology landscape indicators, and EPA River Reach (RF3) data for Region III.

MAIA–Estuaries is developing a standard set of EMAP results that includes cumulative frequency distributions with confidence intervals and estimates of aerial extent of different conditions. This group will also be using and testing a number of environmental indices. They will continue working on the benthic index, and will be developing a fish index and habitat index. They will also use the Landscape Ecology landscape indicators to find associations between landscapes and the estuarine indicators.

Data Distribution
Summary data will be made publicly accessible on the web sites of the organizations that collected or created them (e.g., Chesapeake Bay Program). This policy allows data collectors to maintain their own data and to ensure that duplicate data are not posted. AED will distribute summary data to research partners via anonymous FTP sites and to the public via the EMAP Public Web Site. The EMAP Data Directory will link users these sites. Data will be distributed in a variety of formats, including SAS, Arc/Info export (E00), ASCII (CSV, TXT), and Adobe Acrobat (PDF).

At the end of the MAIA program, AED will transfer the summary data and analytical tools developed to groups in the MAIA region for their use in managing resources. Region III is gearing up to take on the responsibility for the assessment data, the methodology, and the data standards. However, AED is responsible for the long-term maintenance of data it collected.

Research collaborators will maintain the original data on their own servers (i.e., data collected by Chesapeake Bay Program will be accessible on their web site; data collected by AED will be main on EMAP Public Web Site). Access to raw data will be by request.

Data Exchange
This Working Group needs to exchange preliminary data among researchers but cannot use the EPA Intranet for this task because non-EPA researchers (e.g., NOAA) cannot access the network. Preliminary data cannot be posted on the Public Web Site because it has not been through the complete QA process. In the short-term, their goal is to distribute preliminary data to research partners as e-mail attachments in formats compatible with group needs (SAS, CSV, ASCII, etc.). However, improved methods of data sharing must be implemented. This gap is a major deficiency in the current system and needs to be addressed as soon as possible. EPA's investigation of using extranet capabilities on the EMAP Internal Web Site for making preliminary data available to research partners would be beneficial.

Data Documentation
Not all data sets have yet been fully documented. AED is creating EMAP Data Directory and Data Catalog (metadata) entries for data sets it collected. Documentation standards that cover most of the Data Catalog requirements were distributed to collaborators. However, most researchers are still analyzing data, and to date, AED has only received documentation in the form of cruise reports from NOAA. The rest of the documentation will be done some time in the future, and AED is not certain if collaborators will use the EMAP standards.

Data Sources
MAIA–Estuaries will use data from several sources as base maps and additional information, including:

- EMAP Landscape Ecology indicator coverages (clipped to the USGS HUC boundaries) that were developed for the MAIA Landscape Atlas;
- USGS Digital Raster Graphs (DRGs; 42 CDs of elevation data for all the MAIA watersheds) to develop watershed boundaries in ARC GRID for the estuaries they sampled; and
- Region III EPA River Reach (RF3) data (other RF3 data in the MAIA region have not yet been updated, so they will only use this portion).

Data Volumes
Approximately one gigabyte of data could be generated and stored in each monitoring year. Estuaries data currently consists of approximately 20 MB of SAS and Arc/Info files at AED.

Users
In order to justify the expense and maintenance of an EMAP–IM system containing these data, it will be important to understand and serve those who will actually use the data and its documentation.

Primary users of the raw data are the data collectors who conduct analyses and produce summary data sets. Primary users of the summary data include the assessment teams at AED and GED (and their research partners).

CBAT and MAIA will want access to processed data (mostly GIS data) to do their assessments. The majority of assessments will be done by experienced EMAP analysts and research collaborators (e.g., Kent Thornton, Mike Barbour, John Scott).

The Chesapeake Bay Program will also want access to data sampled by other EMAP teams (e.g., AED) in Chesapeake Bay.

Algorithms, Models, Equations, and Other Tools

AED will produce methodology for integrating data from the collaborators in the assessments. The exact approaches to be used are emerging, and ongoing research is addressing many of the issues associated with combining data from a variety of sampling designs and different resources into cross-resource assessments.

Additional research is ongoing on the development of environmental indicators. The indicator development projects are being conducted by EMAP and its research partners.

AED will also conduct comparative assessments of ecological indicators (for example, AED's assessment team is now comparing the EMAP Benthic Index with the Chesapeake Bay Benthic Index on a site-by-site basis). Other indicators will be compared in the future as appropriate questions are posed. Many of these questions are research questions that will be addressed as they arise.

Software, Hardware, Infrastructure

AED is currently using a local area network tied to Microsoft NT and DEC Alpha servers. The network is based on Microsoft NT LAN servers. AED has over 10 gigabytes of disk storage available on the servers and additional storage capacity is added as necessary. Workstations are PC-based, running Windows95.

Project Management

Lack of Adherence to Data Standards

The biggest obstacle to efficient processing and use of MAIA Estuaries data is that research collaborators and contractors are not following EMAP standards and available procedures for data collection (e.g., Oracle Forms entry screens, bar coding procedures), submission (e.g., format, content), and documentation (e.g., EMAP Data Catalog/FGDC). These standards were approved by the research collaborators. Collaborators use their own standards and submit summary data to AED in many different formats. The collaborators may find it more convenient and internally logical to

use their own standards, and they may not be comfortable with using EMAP standards. They may also see the formatting as extra work for which they are not adequately funded.

Lack of standard data formats and documentation makes it more difficult for EMAP to complete the assessments because AED has to re-format and QA the data for its own use. In the long-term, this lack of adherence to standards will cause problems for the EPA regions and state agencies that take over the data from ORD in a few years. The regional agencies might not be able to handle the variety of formats that are coming from the collaborators. Further, this problem of not following standards will continue to be a factor in the CENR framework, in which participants take advantage of data in existing programs, but do not dictate standards to them. The data collection tools, data forms, sample tracking and data submission standards could be a legacy to leave to regional groups for their future work, but it will take increased effort to get them adopted by the participants. However, when the regions take over the data management, there may be even less leverage for requiring adherence to minimal standards. AED will do some outreach to find ways to increase participation in the standards.

Data Delivery by Researchers

Data delivery is more difficult in the current program than it was in the early program because EMAP managers no longer control data collection and management, but instead depend on a network of participating researchers to make their data accessible and understandable. ORD could institute incentives to improve the delivery of data by researchers, including: recognizing data sets and documentation as deliverables that receive professional credit (if this is done, it will be important to ask researchers how they would like their data sets to be cited). The interviewees cited the Human Genome Project's success in data delivery; it accomplishes this goal by not allowing publication of reports in its journal before researchers have submitted their data to the central database.

How Can EMAP–IM Help Them?

The group cited several ways in which EMAP can provide assistance:

- provide EPA extranet or other solution to improve data exchange among research partners
- encourage collaborators to follow standards
- explore idea of ORD giving professional recognition for data delivery and documentation so that it becomes a deliverable along with reports and other products
- maintain the data they collect and create in the long-term and make it accessible on EMAP Public Web Site
- provide a directory and links to data and documentation distributed by research partners
- provide version control for its data and documentation and provide guidance to research partners who want to use these methods

B.4.3 MAIA–Landscape Ecology

Studies being conducted by the EMAP Landscape Ecology Working Group for the MAIA project are summarized in Section B.6, Landscape Ecology. The principal data products are landscape indicator coverages and the MAIA Landscape Atlas, which are now available on the EMAP Public Web Site.

B.5 Intensive/Index Sites

The Intensive/Index Sites Working Group has been subdivided into two main research areas:

- the Demonstration Intensive Sites Program (DISPro) cooperatively with the National Park Service
- the Coastal Intensive Sites Network (CISNet) cooperatively with the National Oceanic and Atmospheric Administration

The following sections summarize the overall Intensive Sites research plans, as well as the plans of DISPro and CISNet.

B.5.1 Intensive/Index Sites Program Overview

Summary of Conference Call
July 24, 1997

Participants

ORD-Gulf Ecology Division
- Kevin Summers, EPA Intensive/Index Sites coordinator

EMAP–IM (AED)
- Stephen Hale, EPA EMAP Information Management

Technology Planning and Management Corporation, ITAS Contractor
- Jeff Rosen
- Dillon Scott

What data collection projects are being conducted within the Intensive/Index Sites group?
There are several types of projects/data:

1. DISPro (Demonstration of Intensive Sites Project) projects at National Parks
   a) UV-B monitoring at selected National Park sites
   b) Routine base monitoring (mostly air quality) at National Park sites

c) Individual research projects at National Park sites through EPA grants to external researchers (wide variety of projects, such as amphibian surveys, nutrient deposition from air effects on local eutrophication, etc.)

2. Additional projects
   a) Monitoring of selected coastal sites with National Oceanic & Atmospheric Administration (NOAA) and National Aeronautics & Space Administration (NASA)

What are the details of the UV-B research?
In the first active DISPro project, EMAP is working with the National Park Service (NPS) to capture new UV-B data. UV-B data includes all information downloaded from Brewer instruments at two types of sites: historical data from primarily urban sites (not included in EMAP) and new data from sites in National Parks. These data will be quality assured/quality controlled and processed into aggregate or summary results.

The National UV Monitoring Center (NUVMC) at the University of Georgia has historically managed the raw data, downloading them from the instruments into a database, running QA/QC checks, backing them up, and distributing some statistics. For new data collection efforts, NERL will put out an RFP for capture, quality assurance, and storage of the new UV-B data. The UV-B data managers will provide the downloaded data to ReVA and EMAP. Aggregate data created during analysis will be managed by EMAP–IM (AED). The ultimate repository of the data is currently under discussion. It may be placed on the EMAP Public Web Site, or it may become the responsibility of NERL and be distributed from their EIMS. NPS is interested in accessing the summarized results, but not in capturing and managing the data (they would like someone else to be the owner of the database).

Both EMAP and ReVA need to use the UV-B data, so it must be accessible to both groups. Documented standards are needed to provide convenient access to the information for both groups. Ongoing negotiations between EMAP–IM and ReVA information management are intended to resolve how the data will be processed and stored and documented. EMAP–IM needs to work with Gary Collins (NERL-Cincinnati) to identify common protocols and formats so that EMAP can access needed information from ReVA, including both the data and documentation. Currently, access to ReVA information systems is password-protected, which limits access by EMAP researchers and managers.

It will not be necessary to distribute the raw data, only the data that has been quality controlled, as well as the aggregates.

Data documentation should include information on how the data were captured, as well as assumptions and indications of specific conditions, the method for developing the index, and the summary results.

AED and NERL (Gary Collins) are working on how to maintain the old UV-B data and what format the data should be in. They are also evaluating which information should be stored online. Currently, it appears that only two variables will be stored (total column ozone and UV-B Index). Historical data for these two parameters only are being loaded by Technology Planning and Management Corporation (North Carolina) into the ReVA Environmental Information Management System (EIMS). Some subset of the data (e.g., total column ozone) is being submitted to the World Ozone Information Center in Toronto (Ontario, Canada), where they are accessible to the general public.

What are the details of the Routine Monitoring at National Park sites?
These are mostly air quality monitoring data. Technology Planning and Management Corporation (TPMC) and EMAP–IM (AED) should talk to John Ray (works for Kathy Tonneson at NPS Air in Denver), who is responsible for all air monitoring data. NPS will maintain these data, and EMAP–IM should get their protocols and QA so DISPro can access it.

What are the details of the DISPro Research Grants data management?
TPMC/EMAP–IM should talk to Bill Hogsett (ORD-Western Ecology Division). EMAP–IM (AED) will handle these data in the EMAP–IM system for conducting research on sampling design, and establishing index sites, monitoring programs, and indicators.

What are the coastal projects in Intensive Sites?
A coastal component is currently being designed for inclusion in the Intensive Sites effort. This component will include approximately 40 sites which will be monitored for some set of parameters. This is a cooperative project between EMAP, NASA, and NOAA. It is anticipated that there will soon be a request for assistance (RFA) to organize support for this effort. The contacts within NOAA are Andy Robertson and Becky Smyth. The contact at NASA is Jim Yoder. Stephen Hale will look into talking with participants about the data (e.g., Kennedy Space Center site).

What will be the Intensive Sites Working Group data needs?
The data needs will vary for each project. Specific plans for how the DISPro data will be integrated and used by EMAP should be addressed to Bill Hogsett in Corvallis, and to the individual research projects. EMAP–IM should also talk to NOAA and NASA contacts about the coastal sites.

What is the Project Management for Intensive Sites?
Each project in Intensive Sites will have its own project management, which will be coordinated by the EMAP Intensive Sites coordinator (currently Bill Hogsett).

How can data submission by individual projects be assured?
Each group that receives funding will have written into the "special conditions" of their award the information requirements specific to their project (it will not be included in the RFA). The data guidelines that EMAP–IM provided to the Regional EMAP program (R-EMAP) could provide a model for these information requirements. However, the requirements will need to be customized for each project and each participant. Requirements should allow principal investigators to have first priority in using the data and include a grace period in which the principal investigator has sole access to the information for publication purposes. In the long-term, these data collected with public monies must be made available to other researchers and the general public.

How are the data currently managed and who are the stewards of the data?
Currently, the data are being managed by the individuals within each project (i.e., University of Georgia manages DISPro's UV-B data). It is not currently clear how the data will be managed for other projects or how long the term of stewardship and maintenance for these data will be. The EMAP–IM needs to identify how data will be managed and how links will be made to data sets being maintained by the principal investigators or other organizations. The long-term disposition of the Intensive Sites data will be investigated by EMAP–IM in further Requirements Analysis interviews with Intensive Sites project managers (Bill Hogsett, John Ray, Jim Yoder). However, it is clear that since these are intensive sites, the long-term maintenance of the data is critical.

NOTES:
It will be necessary to interview Bill Hogsett about the mission/goal statement for the Intensive Sites. Kevin Summers indicated that one of the major objectives of this project was to force cooperation between federal agencies and other organizations in the continued collection of data at Intensive Sites and establishment of long-term trends.

It is expected that there will be DISPro meeting in November 1997 to coordinate ReVA and EMAP efforts. Many of the issues regarding the Intensive Sites may be resolved at this meeting. Stephen Hale will find out if EMAP–IM can talk to participants then about the data.

B.5.2 Demonstration Intensive Sites Project

The Demonstration Intensive Sites Project is a cooperative effort between the National Park Service and ORD's Western Ecology Division.

Summary of Conference Call
April 13, 1998

Participants

National Park Service
- John Ray

ORD-Western Ecology Division
- Bill Hogsett

ORD-Narragansett
- Stephen Hale

Technology Planning and Management Corporation, ITAS Contractor
- Jeff Rosen
- Dillon Scott

DISPro Mission and Goals

The Demonstration Intensive Sites Project (DISPro) will conduct monitoring at selected National Park sites where air monitoring is ongoing. This project represents an inter-agency effort between EPA/ORD and DOI/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." Twelve national parks were selected according to selection criteria (e.g., readily accessible, history of monitoring data, contain a broad spectrum of ecological communities).

DISPro will focus on long-term effects research for atmospheric data (UV-B and air quality at sites in the Parks that already have 10–15 years of data for comparison, such as air deposition). The intent of the program is to initiate a consistent air monitoring program at each site to be followed by monitoring within other media. In order to demonstrate the relevance of this monitoring, research projects will eventually be initiated at all of the sites to examine the effects of environmental stressors of importance.

General information about ongoing NPS air monitoring programs is available on the NPS web site (NPS 1998a), plus information on the Air Resource Division web site (NPS 1998b) and the Inventory and Monitoring program web site (NPS 1998d).

DISPro includes the following components: UV-B monitoring, routine NPS air quality monitoring, and individual air quality research grants. The first project that will be conducted is the UV-B

monitoring and will be reviewed in the following subsection. The other projects are new to EMAP and more information will be added as it becomes available. These projects are summarized below.

The DISPro Home Page on the EMAP web site may consist of links to other sites where data actually reside (e.g., EIMS, WOUDC, CASTNet).

B.5.2.1 UV-B Monitoring

The purpose of data collection in this group is to develop long-term monitoring databases for the sites. The initial focus is on UV-B, but they also collect other spectral bands and UV-A that could be used more extensively in the future.

Data Collection, QA/QC, Analysis, Aggregation

The purpose of collecting UV-B data is to measure full-sky solar UV-B and UV-A spectral flux, from which absolute irradiance and total column ozone concentrations are calculated. The data are intended for dissemination to government and non-government scientists and interested parties. The effort also collects some routine air monitoring parameters (e.g., wet or dry nitrogen, ozone).

The National UV Monitoring Center (NUVMC) at the University of Georgia (NUVMC 1998) manages the UV-B measurements from high spectral resolution spectroradiometers at 12 NPS monitoring sites. The NUVMC is part of the UGA/EPA UV Monitoring Network (UVMN), which operates and maintains a group of high spectral resolution spectroradiometers throughout the United States. The NUVMC downloads data from the collection sites each night 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 two kinds of data files: raw data, and data that has been through first-level error checking. NUVMC also produces plots of total UV (DUV/irradiance) and total column ozone and places the plots on its web site. The EPA Environmental Information Management System (REIMS) will provide data management support.

Data Distribution

The University of Georgia archives the processed raw UV-B data in a database. They also produce DUV and ozone plots that are made available to users on a University of Georgia web site (NUVMC 1998); electronic files of the processed raw data can be obtained by request.

Long-term archiving of summary data will primarily be in the EIMS database. ReVA is now determining a subset of parameters that will be made available on the EIMS web site. The site will provide access to the full data set only for authorized users, not for secondary users. The EIMS web site will become the authoritative site where EMAP and NPS researchers store their improvements to the data as they use it in analyses.

ReVA will send a QA/QC'd version of the raw data to the Canadian "World Ozone and UV Radiation Data Centre" (WOUDC 1998) for distribution to secondary users. The WOUDC is one of seven recognized World Data Centres that are part of the World Meteorological Organization (WMO) Global Atmosphere Watch (GAW) program. WOUDC will apply data validation and proofing techniques that they use on their extensive data holdings. Archived data will be available by conventional FTP access (password required) and by a direct link from the WOUDC Data web site. The data will be redundant with that storedin EIMS, although the EIMS version will not have been processed with the WOUDC validation and proofing techniques. The EIMS web site will be the preferred site for NPS and EMAP users because metadata and iterative improvements to the data will be maintained there.

Data Integration

The current obstacles to integrating these data into EMAP assessments include:

- The data are not currently available in formats that facilitate consistent integration among air measurements from the different researchers. Researchers must individually reconcile data sets. The data must be made available for consistent integration among the sites, using consistent interpolation techniques (e.g., GIS). for calculating exposures across the parks and building maps of park species cover, location of research plots, exposure patterns. These interpretations would be given to the Parks to give them an indication of where the past three years of surveys had occurred for use in planning future surveys (Parks need a way to keep consistent information from one set of researchers to the next, with a loss of opportunity to sample in the same areas); and
- Researchers cannot easily access the hourly data, which is needed for response models and statistical analyses (yearly averages available through AIRS are insufficient).

Data Documentation

UV-B data will be documented by the field support contractor. Documentation must include site information (log sheets, audits, calibration records, maintenance records). The documentation will reside on the EIMS web site. Methodology and data reduction SOPs should be written by the contractor. The NUVMC already has something like this but more detail could be added (project officer must decide what level is needed).

Data Sources

The Intensive Sites Working Group is not using any external data sources except that in the future, they will use some of the data NPS has previously collected at the intensive sites.

Data Volumes

Large volumes (amounts unknown) of instrument data and data aggregates will be stored in the EIMS database.

GIS/Georeferencing
Latitude and longitude is recorded in the database for all data collection sites. The format of the latitude and longitude is in different formats in different databases and will have to be reconciled for use in assessments. There has not been any specific coordination on location data among the different data collection groups. This is an important issue that will be addressed by the Working Group in the future.

Users
There is not yet an established user community for these data except those directly involved in the research, including:

- Brewer maintenance staff in parks;
- NUVMC University of Georgia; and
- EMAP and NPS UVB researchers (these researchers understand the instrument and can process the data and know what to do with it).

There is no specific plan for who the users will be and what part of the database they will want to use. The current purpose is to collect UV-B data—which is difficult to collect and not widely available from other sources—under an available opportunity so that it will be stored for future projects. Potential users include:

- EMAP routine air monitoring researchers
- EMAP air monitoring grants researchers—There will be a need for most researchers doing effects studies under EMAP air monitoring grants in the Parks to compare their data with the DISPro data. To do this, they will need to access to the exposure data being collected along with the UV-B data (e.g., wet or dry nitrogen, ozone). Currently, arrangements will must be made for them to access needed data.
- EMAP Regional Assessments researchers
- EMAP assessment researchers and management

Secondary users may include those who now access the NUVMC and WOUDC sites.

Software, Hardware, Infrastructure
The equipment and software for managing these data will be taken care of at individual research and data repository sites.

How Can EMAP–IM and the EMAP–IM System Help?
The interviewees indicated that the most useful thing will be to make the EMAP–IM system as user-friendly as possible. They would like to have EMAP–IM assistance with creating EMAP Data Directory entries for UV-B data. They do not need assistance with routine data collection and management, or a data repository. They would like to have links from the EMAP Public Web Site to their data repository locations.

Interaction with Other Working Groups
Intensive Sites will interact with the Regional 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; if the next assessment is out west, there will be a greater number of parks in the area, which will potentially cover Regions VIII, IX, and X.

B.5.2.2 Other DISPro Initiatives

Routine Air Quality Monitoring
This program currently has 31 sites (including the 14 DISPro sites discussed above). A contractor (Air Resource Specialists/ARS) provides field support and data handling; checks stations every night; puts data into an Oracle database; does four levels of data validation; and archives the data from each level of validation. The final data set is stored in an Oracle database at ARS. ARS is working on setting up Oracle database for Internet query; it is at the prototype and testing stage and will eventually be open to the general public. NPS submits all ozone, sulfur dioxide, and meteorological data to the EPA AIRS database (AIRS 1998), which is fully accessible only to users with accounts. Data submitted to the AIRS web site is currently only available in annual summaries, but the one-hourly data that NPS submits to AIRS will be made available within a year.

Individual Research Grants
Individual external researchers will be funded to conduct intensive sites monitoring on a wide variety of topics (e.g., amphibian surveys, nutrient deposition from air effects on local eutrophication). EMAP–IM will handle the data and enter it into an Oracle database for use in future sampling design and other research on setting up index sites, monitoring programs, and indicators.

B.5.2.3 Related NPS Data

EPA Clean Air Status and Trends Network (CASTNet)/National Dry Deposition Network (NDDN) Monitoring Results of Emission Reductions
These data include dry deposition filter pack data, which can be accessed on the EPA CASTNet web site (CASTNet 1998) and the NPS Dry Deposition Monitoring Network (NPS 1998b). In order to access these data, it is necessary to pay for an account on the system.

Wet Deposition Data
NPS wet deposition data collected at the sites may be placed on the CASTNet 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 a web site at the University of Illinois (UIUC 1).

IMPROVE Filter Pack Data
This is a joint EPA-NPS-U.S. Forest Service (USFS) program to study visibility in the Parks. The prime contractor (University of California, Davis) makes the data available from an anonymous FTP site for all the years of data collection. The site has ASCII files of data by individual site location. It also has five particle measurements, four modules with filter packs that collect different things. There are other data not currently on the FTP site that must be requested directly from ARS, including optical measurements with nephelometers, transmissometers, and other equipment.

NPS Inventory Monitoring Program
This is biological survey data collected in the parks (e.g., amphibians). EMAP will provide common links to these data in the future.

B.5.3 Coastal Intensive Sites Network

No interview was conducted. See the summary in Section 2.3.2.2 (CISNet), which was prepared from information at the NCERQA web site (CISNet 1998).

B.6 Landscape Ecology

Landscape Ecology Working Group is based at the NERL lab but has cooperators at many sites (e.g., U.S. Geological Survey). The interview was conducted with NERL staff.

Summary of Workshop
September 24-25, 1997

Participants

U.S. EPA Las Vegas Laboratory (NERL)

- Deb Chaloud (702)798-2333 chaloud.deborah@epamail.epa.gov
- Curtis M. Edmonds (702)798-2264 edmonds.curtis@epamail.epa.gov
- Sue Franson (702)798-2213 franson.sue@epamail.epa.gov
- Ed Furtaw (HERB-LV) (702)798-2285 furtaw.ed@epamail.epa.gov
- Daniel T. Heggem (702)798-2278 heggem.daniel@epamail.epa.gov
- Stephen Hern (702)798-2594 hern.stephen@epamail.epa.gov

- Bruce Jones (702)798-2671 jones.bruce@epamail.epa.gov
- Bill Kepner (702)798-2193 kepner.william@epamail.epa.gov
- Clay Lake (702)798-2269 lake.clayton@epamail.epa.gov
- Bob Schonbrod (702)798-2229 schonbrod.robert@epamail.epa.gov

EMAP–IM (AED)

- Stephen Hale (401)782-3048 hale.stephen@epamail.epa.gov
- Melissa M. Hughes (401)782-3184 hughes.melissa@epamail.epa.gov (OAO Corp.)

Technology Planning & Management Corporation (ITAS contractor)

- Jeff Rosen (781)544-3085 jrosen@tpmcscituate.com
- Dillon Scott (781)544-1298 dscott@tpmcscituate.com

Roles of the participants are as follows: Bruce Jones is the overall lead for the EMAP projects and the EMAP Landscape Ecology Working Group; Robert Schonbrod is the Branch Chief for Landscape Ecology; all other Las Vegas staff are managers of individual EMAP projects.

Purpose of the Meeting

The purpose of this meeting was to:

- establish the functional and structural requirements of information processing for the Landscape Ecology Working Group; and
- determine the appropriateness of the Landscape Ecology requirements and the IM planning being performed by EMAP, and the information management relationship between the two programs.

Landscape Ecology Mission

The mission of the Landscape Ecology group is to integrate data from multiple scales (remote sensing images and site data) to develop landscape indicators and assessment protocols for EMAP and ReVA. Landscape indicators will be used to evaluate landscape status from 1992–1993 MRLC imagery. Landscape Ecology will use these data and gradient studies to develop indicators. They will evaluate the impact of landscape change by trading space (spatial variability—i.e., a gradient of spatial variability) for time. A few indicators and assessments will involve landscape change over a 20–year period.

For EMAP, the group will develop landscape indicators and assessments of status and trends for selected resources of human importance (e.g., water quality, habitat quality). For ReVA, they will use the indicators to evaluate or assess resources for their potential vulnerability to future degradation as a result of multiple stressors (for selected resources of human importance (e.g., water quality, habitat quality). 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. The current ReVA projects focus on streams and forests. The current Landscape Ecology projects for EMAP include those listed in the next subsection, **Studies.**

Studies

Landscape Ecology is conducting two main projects for EMAP:

- Mid-Atlantic region (MAIA (U.S. EPA 1995d); and
- R-EMAP projects (in cooperation with EPA Regions; data from these projects will be submitted to the Regions)
  - ❑ Region IV—Savannah River Landscape Analysis (lead: Deb Chaloud)
  - ❑ Region VII—Landscape Analysis and Characterization to Support Regional Environmental Assessment Project (lead: Bruce Jones)
  - ❑ Region VIII—Integration of Upland and Riparian Stream Condition Monitoring for Intermediately Sized Watersheds on Rangelands (leads: Anne Neale, Dan Heggem)
  - ❑ Region IX—Bioassessment of Water Quality in the Humboldt River, Nevada (leads: Anne Neale, Dan Heggem).

Related projects for ReVA include:

- National Water Assessment (for Office of Water); data now listed on EPA Surf Your Watershed web site (EPA 1);
- San Pedro Basin (Arizona—Mexico) (ReVA 1998a, ReVA 1998b);
- Tensas River, Louisiana; and
- Lower Colorado River (LoCo) (planned).

Users

Landscape Ecology users include a number of research collaborators who exchange data at all stages of the work, and a number of end users who receive data after QA/QC, aggregation, and analysis are completed.

Primary users include:

- EMAP and ReVA scientists and analysts;
- planners at all scales (local, regional, state, federal); and
- resource managers at all scales (local, regional, state, federal).

Secondary users include:

- educators;
- universities;
- public (including insurance, lawyers, media, etc.); and
- international agencies (e.g., World Bank, Mexico).

Landscape Ecology indicated that there is a critical need to support increased user requests. In particular, they need better data distribution mechanisms (e.g., public FTP, staff support or clearinghouse) to collaborators and end users. The **Issues** subsection, below, contains more information. Interviewees also stressed access for educational users because Landscape Ecology data products (e.g., Mid-Atlantic Landscape Atlas) are valuable outreach tools, and Landscape Ecology staff have presented materials to schools in the past. Such materials and data could be distributed to EPA educational programs and schools on CD-ROM and via public FTP, but currently EPA does not allow this type of access.

Data Sources

Landscape Ecology uses data from a wide variety of local, regional, national, and international sources, including:

- Satellite imagery, such as:
  - ❑ MRLC 1992 nationwide coverage (obtained from RTP Landscape Characterization group)
  - ❑ NALC data (Landsat MSS) from 1970s, 1980s, 1990s (Landscape Ecology is in charge of this database in cooperation with USGS). Currently, there is more than 0.25 terabyte (TB) of this data for the lower 48 states and Mexico (an additional 0.10 TB for Alaska and Hawaii if added to the database). The data are currently being converted to FGDC standard format.
  - ❑ AVHRR for analyses at coarse scales;
  - ❑ Primary site data (locations, measured values), such as:
  - ❑ MRLC land cover
  - ❑ USGS 90-m DEM
  - ❑ STATSGO soils (NRCS 1:250,000 state soil data)
  - ❑ USGS DLG roads
  - ❑ U.S. Census
  - ❑ EPA atmospheric deposition
  - ❑ EPA River Reach RF3
  - ❑ USGS HUC 250; and

- Early EMAP data (especially Surface Waters, Forests, Agroecosystems, Arid or Rangelands, R-EMAP).

In the future, Landscape Ecology will also need access to:

- data from new satellite sensors;
- NDVI analyses/land cover;
- new data layers from other agencies and independent sources (e.g., soil);
- combination of satellite imagery with radar;
- data from modernized STORET;
- distribution of historical and existing data sets on CD-ROM (e.g., Nature Conservancy);
- new process models, algorithms, etc. (need a metadata directory of them); and
- new EMAP data: Multi-Tier Design, Intensive Sites, and Regional Monitoring Pilots or Demonstrations.

The Landscape Ecology group also identified some important data gaps that need to be filled for them to conduct their research, including:

- large-scale wetland data;
- large-scale soil maps (especially of texture);
- forests;
- geology;
- finer-resolution air quality data;
- road networks;
- 11-digit USGS HUCs;
- river flow rates;
- DEMs;
- wetlands;
- flood plain maps;
- SO_2:NO_2 at elevation;
- 1970s NALC imagery (lacks completeness of coverage); and

- water quality (USGS gauging station locations do not integrate watershed characteristics).

The source agencies for some of these data include:

- USGS Earth Resources Observing System (EROS);
- Universidad Nacional Autonoma de Mexico (UNAM);
- U.S. Department of Agriculture (USDA);
- USDA Agricultural Research Service (ARS);
- Instituto del Medio Ambiente y el Desarrollo Sustenable del Estado de Sonora (IMADES); and
- Arizona State Land Department (Arizona Land and Resources Information System).

The group indicated that it is particularly important for them to have access to early EMAP data sets in order to validate landscape indicators and assessments. However, they are currently having difficulty accessing data such as Agroecosystems and Forests in a format and resolution appropriate to their needs.

The quality of documentation of data sources varies widely. Less than half of all data sources are not well documented or organized, and it can be expensive and time-consuming to determine the suitability of data for analysis and integration. Landscape Ecology indicated that they would like to be able to track when and how data sources have been updated or corrected, so they can determine how the changes affect previous or future analyses.

There is also a lack of consistent quality and completeness of data sources. It is often difficult to obtain data of sufficient quality for landscape analyses, or they can only get data for part of a geographic area (e.g., for EPA River Reach data, only the Region III updated portion was detailed enough to be used in their Mid-Atlantic project).

The longevity and quality of data stewardship also varies, and Landscape Ecology must depend on source agencies to maintain and make their data available.

The group agreed that the EMAP Data Directory would be very useful for solving some of these problems of locating data sources and documentation. Locating data sources is a time-consuming task, especially for local, regional, and independent (e.g., private, such as the more than 300 aerial photography vendors across the country) data sets. Such local sources are the most difficult to find, and usually it is necessary to invest a large amount of time to locate them. One suggestion made was

that each time EMAP starts a new project, a page could be initiated on the EMAP web site that would allow all participants to post and share data sources.

Data Management

Data management for Landscape Ecology is currently accomplished by the researchers, who do their own data entry, data QA/QC, data aggregation, data analysis, data distribution, and system maintenance. Some support is also available from on-site contractors. Currently, they do not have dedicated data management personnel who can support secondary data distribution, detailed documentation, and widely accessible data archives. This arrangement successfully supports research objectives, but does not provide sufficient resources to support distribution of data to EMAP and other user communities (see **Data Distribution** and **Documentation** subsections, below).

Data management tools used are primarily those in the existing software (e.g., ARC/INFO, dBase). They are planning to increase their use of Oracle RDBMS tools in the future.

The group indicated that it would be valuable to have assistance and resources for coordinating and documenting their data management procedures—which currently vary for each project's special needs. These resources would be helpful with: 1) providing data management training to develop Standard Operating Procedures (SOPs) for capture, storage, and dissemination of data; and 2) training in the use and optimization of relational database management tools (e.g., Oracle).

Landscape Ecology will need to add data management staff in order to address these problems of data management and distribution (see **Issues** subsection, below)

Data include the imagery and tabular data sets, surface maps, ARC/INFO coverages, metadata, and processing algorithms, procedures, and programs. While it is not possible to estimate the exact volume of each data set individually, Landscape Ecology estimates that it will be necessary to store a total of about 0.5 terabytes +/- 0.25 (TB) per year. Currently, the database size is approximately 250 gigabytes (GB).

Data flow now includes:

- data collection—determining data needs, obtaining data from sources, collecting new field data;
- data entry/conversion—transferring data into a form compatible with analyses and major data sources (e.g., converting analog to digital by digitizing; re-projecting spatial coordinates). SOPs exist for some procedures (e.g., UTM projection to latitude/longitude);
- data verification—conducting range checks and other minimum quality checks;

- data validation—is performed while using the data (e.g., checking dates and names, and using software which checks an image from a CD-ROM);
- data QA/QC—checking data to determine suitability for analysis/integration; SOPs exist for some procedures;
- data analysis—using appropriate procedures/software to create indicators, conduct status and trends, and assess vulnerability; SOPs exist for some procedures;
- data aggregation—putting data in common projections, correcting mistakes, etc. (see **Data Aggregation** subsection, below);
- data distribution—transferring indicator data, atlas images, etc. to users (see **Data Distribution** subsection, below);
- data storage—storing images and data on 8-mm tape archives to CD-ROM; currently moving archives to CD-ROM; and
- data security—there is a need to control access to data sets under development.

There can be loops in this data flow; for example, when data inadequacies are found during data analysis, it may be necessary to return to data collection or entry.

Data Quality Assurance/Quality Control

The ultimate responsibility for Landscape Ecology data quality and maintenance resides with the individual project manager, as well as the working group leader, Bruce Jones (see **Project Management** subsection for more information).

Landscape Ecology has SOPs for some QA/QC procedures, principally for digitizing and data processing. They use standard existing datums and projections (e.g., NAD83) for spatial data.

Data Aggregation

Data aggregates are the major data product from Landscape Ecology, and consist of value-added source data, (e.g., co-registered data sets, corrected images), as well as new data from analyses. Data aggregates contain information that is frequently requested by other users. For example, Landscape Ecology adds value to many of the data sources they use by synthesizing, correcting, and augmenting them for their analyses. Their integration of early EMAP data sets with 1990–1993 satellite imagery will produce aggregates of site data and surface maps that will be valuable for many future projects. These aggregates represent new data sets, distinct from the original source, that must be documented and cited in the Data Directory.

Data Products

Landscape Ecology creates a variety of data products and tools from their research that are valuable to other researchers for site studies, regional/national assessments, status and trends, education and outreach, and other tasks. These products will be produced in hard copy and electronic form, and include:

Table B.2 Data Products

Models, methods, protocols, algorithms	electronic // paper	spreadsheets, programs // publications
Programs	electronic	C++, etc.
Surface maps (co-registered with other spatial data)	electronic // paper	remote sensing images, ARC/INFO coverages, ASCII files // Atlas maps
Linked GIS data coverages	electronic	ARC/INFO coverages
Assessments (output of analyses)	electronic // paper	ARC/INFO coverages, spreadsheets // Atlases, publications
Source data sets (modified for use in landscape analysis)	electronic	spreadsheets, ASCII, ARC/INFO
Regional & special studies results	electronic // paper	ARC/INFO coverages, spreadsheets // Atlases, publications
Model output/runs stored as primary data, model (algorithm), and output (grid)	electronic	spreadsheets, ARC/INFO coverages, ASCII files
Metadata	electronic // paper	ASCII and WP files, publications
Publications	electronic // paper	Atlases, scientific journals, symposia proceedings
Presentations to educational and policy groups	electronic // paper	ArcView Project Files, publications, Fact Sheets, Videos
Atlases	electronic // paper	Acrobat Page Maker Files, Publications
CD-ROMs (containing all of the above products)	electronic	Acrobat Page Maker Files, ARC/Info Grid and Coverage Files, ARC/INFO Export Files (.E00), ARCVIEW production files

Atlases are the main published products and include indicators mapped at multiple scales. An Atlas of Chesapeake Bay has been completed (Riitters & Wickham, TVA & EPA, 1995) and contains land cover data, indicators, and landscape assessments. The Region III portion of the Mid-Atlantic study has been produced as an Atlas and is now undergoing EPA pre-publication review; it will be made available in printed and online formats. The online version of the Mid-Atlantic atlas is expected to

be greater than 1 and less than 2 GB (250 images at ~2–30 MB for each image; each Atlas page is approximately 30–40 MB), and would fill approximately 2–3 CD-ROMs.

Landscape Ecology is also planning to produce comparisons of landscape indicators and ecological indicators from other Working Groups.

Data Exchange Among Research Partners

Landscape Ecology particularly needs to frequently exchange preliminary versions of large geographic data sets (2–500 megabytes) with co-researchers who share primary responsibility for new data, data aggregates, data analyses, and methodology on the project (partners include two other EPA programs located at different sites (Landscape Characterization at RTP and USGS Reston), as well as the Tennessee Valley Authority (TVA), MRLC and NALC participants, Department of Energy Oak Ridge National Lab, The Nature Conservancy, U.S. Forest Service, and New Mexico State University). The data sets must be made available to co-researchers in a short time frame (less than one week) on an ad hoc basis. However, there is currently a bottleneck in this process because large file sizes are often too large to transfer via email and traditional transfer media, and EPA policy does not allow Internet access to EPA computers other than the public access server at RTP. However, the data cannot be placed there because access to data under development must be restricted to authorized users. The interviewees indicated that it is not adequate to have close collaborators submitting requests and receiving data a week later. The problem has been temporarily address by an FTP site at TVA, but EPA must find a more long-term solution. It is important to improve this group's ability to share data with their partners, cooperators, and other groups by providing write-access to an site where they can rapidly post large data sets (e.g., FTP).

Data Distribution

Distributing completed summary data to end users should be a priority for EMAP and other users because the group produces a large amount of value-added data (e.g., site data mapped onto imagery). The data are useful to all those listed in the **Users** subsection, and especially to research partners and other Working Groups. Providing an efficient mechanism for distributing the data will increase their value to EMAP and the user community.

Landscape Ecology receives many requests for their data aggregates, analyses, and models from non-EMAP research groups and the public. Many of their completed summary data sets (e.g., indicator data) will be posted on the EMAP Public Web Site with links to ReVA Home Page and the EPA Surf Your Watershed web site (SURF 1998), where data will be organized by USGS 8-digit HUC (watershed) codes and other assessment units. The Mid-Atlantic Landscape Indicators Atlas will also be published in the next few months in paper and electronic form, and they anticipate that this will generate a large number of data requests. Although these data can be published at several resolutions (thumbnail, mid-resolution, high resolution), users will still request customization or subsets of the data. Landscape Ecology already does not have sufficient staff to fill existing data requests, so the group discussed options for public data distribution, including:

- create an EPA data download site with interfaces that allow the user to customize their data requests
- designate a data clearinghouse that will fill requests
- privatize the data distribution
- create a data management position dedicated to maintaining monitoring data sets or filling requests

The workshop participants mentioned two national monitoring networks as models for data distribution:

- UK Environmental Change Network (UK Environmental Change Network 1998);
- ERIN, Australia's national monitoring data network (ERIN 1998).

Data distribution formats currently include ARC/INFO GRID export format (.E00) and spreadsheets.

The landscape indicators and assessments may also be used in Surface Water Assessment Programs (SWAPS) under the Safe Drinking Water Act.

Documentation

The Landscape Ecology staff stressed the need for adequate documentation to be distributed with and linked to each data product. Because there are many ways to do the same analyses (e.g., NDVI), it is critical—for legal and scientific purposes—to have good descriptions of data quality and analytical procedures. Documentation should include how source data (e.g., sampling results) were filtered for use in landscape assessments. The group receives many requests for their data from researchers and the public, or results of analyses can end up in litigation; therefore, they feel it is important to provide users with methods and revisions so it can be properly used and understood.

Currently, their existing documentation fulfills minimum requirements outlined in their 1995 Mid-Atlantic work plan, and they are working towards meeting FGDC standards. They indicated that they will need additional resources for completing Data Directory and Catalog entries (this assistance may be provided by ReVA). They would also like to receive guidance on meeting FGDC standards. EMAP–IM can provide examples of metadata templates and data entry tools from EMAP, NOAA/CSC, and other sources.

Once standard documentation is made available, it will be useful for understanding data sets but also for cataloging and accessing needed data. The FGDC documentation format will fit well into the EMAP Data Directory and Catalog model, and FGDC documentation files could easily be linked or converted to the Data Directory and Catalog.

Software
Landscape Ecology uses both standard and custom software tools to process and analyze data, including:

Standard

- ARC/INFO
- ERDAS
- SAS
- S +
- ENVI
- PCI
- ArcView
- Oracle
- Adobe Freelance
- Adobe Acrobat
- WordPerfect
- QuattroPro
- PKZIP and GZIP

Custom

- Programs in C and C++ written in- house and by cooperators
- LandStat/Spatial Convolution
- Risk Assessment Management Software (RAMAS)

Hardware and Infrastructure
The Landscape Ecology group currently uses a combination of UNIX and EPA DECs, but they plan to migrate most of their processing to Windows NT systems in the near future. They will retain a UNIX network server, and NT systems will take on more of the data processing burden. Windows NT is viewed as a less costly, more flexible, and more user-friendly platform with adequate capacity to support their analytical needs.

The group indicated a need for more storage capacity, especially for processed remote sensing images. They regularly run out of disk space for these large data sets (there is now about 250 MB of data on the server, and they will need about 0.5 terabytes +/- 0.25 (TB) per year of additional storage space). Currently, they have to store much of the data off-line on CD-ROMs and 8mm tapes.

They also need broader Internet bandwidth and an anonymous FTP site to handle distribution of their large data files (see **Data Distribution** and **Issues** subsections for more information).

Project Management
Responsibility for each Landscape Ecology EMAP project lies with an individual staff member, who conducts the project and manages the data. The lead responsibility for all projects belongs to Bruce

Jones, and ultimately Landscape Ecology Branch Chief Robert Schonbrod. Bruce Jones is also the official EMAP Landscape Ecology Working Group lead, which means that the group reports directly to the EMAP Director. The liaison between Landscape Ecology and EMAP–IM is Anne Neale.

The Landscape Ecology group is in NERL, but it has a commitment to complete its EMAP projects through an explicit agreement with NHEERL.

In order to exchange data with other working groups and researchers, and to deliver data and documentation to EMAP, Landscape Ecology relies on the same project management tools as the rest of EMAP, including:

- **Common goals**—success of data exchange between the ReVA and EMAP projects will be encouraged by the overlap of data needs between these two programs. Closer ties between NERL and NHEERL can only help this cause. Landscape Ecology also indicated that there may soon be an across-ORD Landscape studies team, organized by Steve Paulsen, which would provide some oversight;
- **Planning tools**—Landscape Ecology's commitments to EMAP are outlined in the NERL-NHEERL agreement and are being monitored through ORD's project management tools OMIS and GPRA;
- **Good will**—data and resource sharing between ReVA and EMAP now depend largely on good will between cooperators in the matrix management setting of EPA; and
- **Funding agreements**—funding is also a major factor in ensuring project success, since NERL commitments to EMAP are explicit in the existing agreement.

Issues

A number of resource and data themes emerged as important issues that need to be resolved for Landscape Ecology to complete their work and make data accessible, including:

1. <u>1990–1995 EMAP Data Sets</u>

   Landscape Ecology needs access to site data that it has been unable to obtain, including:

   - EMAP Forests;
   - EMAP Agroecosystems;
   - several R-EMAP projects; and
   - EMAP Surface Waters data collected after 1995.

2. Data Distribution and System Maintenance Support
In order to fulfill the demand for their data, manage the data, and do system maintenance, Landscape Ecology needs increased staffing support in the form of 2–3 FTE for a system manager, data librarian, and database programmer. The data librarian would be 1 FTE and is needed to support creation and updating of documentation, tracking of data sets, and distribution of Landscape Ecology data sets. The system manager would be 0.5–1 FTE and is needed to support system maintenance. The database programmer would be 0.5–1 FTE and is needed to support the database management needs outlined in the **Data Management** subsection and item 3, below.

3. Data Management and Documentation Guidance and Resources
Landscape Ecology requested guidance and training in database management to support distribution of data, procedures, and results, and development of acceptable documentation that is compliant with EMAP and FGDC standards. Landscape Ecology currently needs additional resources in their budget to develop documentation. The data management tools now being used for spatial data access and analysis (e.g., ARC/INFO, ERDAS) are sufficient for technical personnel to manage data for their own research purposes but not for non-technical users. Support is needed to develop tools for notifying users of data availability, content, status, and updates to data sets (version control). Managing multiple versions of data sets is a major data management and documentation issue that will require application of standard version control software and expertise currently not available in this Working Group.

4. Data Quality
The group's primary mission is to develop landscape indicators and assessment protocols. However, resources are needed to develop and implement a program to validate the indicators and report the accuracy and variability of the validation.

5. Data Distribution
As discussed in the **Data Exchange** subsection above, the group needs access to an efficient method for exchanging their large data sets in a short time frame. EMAP must find a site for this exchange.

6. Data Directory Coordination between EMAP, ReVA, and NCEA
The upcoming pilot project between EMAP, ReVA, and NCEA to integrate the Data Directory for both programs simplify access to information for Landscape Ecology and other users.

Incentives

The workshop also focused on what incentives might improve sharing data and completing documentation throughout EMAP. A number of ideas were discussed that might be helpful in facilitating data access:

- Data sets should be referenced deliverables or milestones that are listed as accomplishments of a project. This feature is already in the GPRA tracking system and should be added to the appropriate planning documents. If this idea is implemented, it may be a good idea to conduct peer review of data sets before they are published.
- Synergy between EMAP and ReVA should be strengthened because common research objectives encourage data sharing, accessibility, and utility.
- Additional staff should be allocated as discussed in the **Issues** subsection to support database management, distribution of data and documentation, and system maintenance for this Working Group.
- The EMAP Director could target specific dates for loading of Resource Group site data to ensure its availability for Landscape Ecology analyses.

How Can EMAP–IM Help Landscape Ecology?

EMAP–IM can offer assistance to help resolve some of the issues raised in the workshop, particularly in the areas of locating and documenting data sources and data products, facilitating data exchanges, providing access to infrastructure, and providing guidance and standards for documentation and data management. Some examples specific to Landscape Ecology include:

- Facilitate access to key EMAP data sets needed by Landscape Ecology (e.g., Forests, Agroecosystems, Surface Waters, etc.);
- Locate, organize, and facilitate access to a wide variety of data and metadata not currently indexed to support Landscape Ecology's work validating indicators and assessments. This access must include an indication of the content and quality of the data sets. Include a broad range of external data sources as Internet links and documentation (data sets of interest to Landscape Ecology cover a wide range of topics mentioned in the **Data Sources** subsection; for example, some additional data sets of interest to Landscape Ecology would include Quonset Hut locations, pesticide use, transportation, and finer scale data on land uses.);
- Provide guidance on preparing Data Directory and Catalog entries by providing examples of existing metadata entry tools and templates;
- Provide site data handling protocols for Landscape Ecology's field data collection based on EMAP 1990–1995 field efforts (e.g., is there a role for a "Lessons Learned" workshop

or publication on SOPs/protocols of successful approaches used in the past for data management); guidance on use of log sheets vs. automated data collection systems;

- Provide an Internet-accessible site where Landscape Ecology data sets, protocols, etc. could be rapidly posted for access by research partners;
- Post guidance for the following topics on the EMAP Public Web Site:
  - ❑ codes for QA, taxonomy, chemistry, etc.
  - ❑ metadata creation tools/templates
  - ❑ tools and lessons learned from 1990–1995 program;
- Keep the EMAP Bibliography up to date so that Working Groups can find publications useful to them. In turn, Landscape Ecology and all Working Groups should send their publication information to EMAP–IM promptly;
- Facilitate an EMAP publications clearinghouse so that different programs can find and access useful publications;
- Assist with making the Mid-Atlantic Landscape Atlas accessible online;
- Put the data on the EMAP Public Web Site and provide linkages to ReVA and Surf Your Watershed web sites; and
- Support access to Landscape Ecology data and metadata.

Conclusions and Recommendations

Our discussions with Landscape Ecology revealed a dynamic group that is producing a large amount of data and processing tools that are useful to many EMAP and outside users. This group is very interested in participating in the EMAP/ReVA data documentation and distribution system, and it will be a worthwhile investment to assist them with making their data and information more accessible via the EMAP Directory and Catalog, and data download sites.

In summary, the primary needs expressed by the group included the following:

- Training and assistance with development of FGDC-compliant data documentation;
- Identification of additional resources for creating EMAP Data Directory and Catalog entries that link the data with the users;
- Guidelines and policies on data documentation from EMAP–IM;
- An Internet-accessible site where Landscape Ecology and other EMAP data suppliers can post large data sets for immediate access by researchers not at the Las Vegas site or with EPA;

- Access to 1990–1995 data sets and new Surface Water data for regional analyses (current and retrospective);
- A publications clearinghouse within EMAP that can distribute publications from the different programs;
- Assistance and/or guidance with management of (potentially) terabytes of raw and aggregated data and imagery;
- Access to raw and spectral data for determining data usefulness to more directly measure ecological processes (e.g., productivity derived from leaf cover index or estimates of the Normalized Difference Vegetation Index (NDVI));
- Assistance with identifying optimal tools for acquiring, managing, analyzing, and documenting environmental information; and
- Distribution of their models, procedure, and algorithms, and access to those produced by other programs.

It is clear from this meeting that the EMAP Data Directory and Catalog approach could be useful to this group for:

- organizing and summarizing programs and data sets so they are easy to locate and understand;
- making metadata accessible to users;
- notifying users of data set corrections and updates;
- linking to existing data web sites;
- working in consortia, to locate contacts, areal coverage of data, and data updates;
- facilitating inquiries about 1990–1995 data sets; and
- listing publications.

All of the participants agreed that it is important for EMAP–IM (AED) and Landscape Ecology to get started entering Landscape Ecology data and metadata into the EMAP Data Directory, Data Catalog, Home Page, and Bibliography. These products will help solve some of the problems with locating, updating, and distributing data.

B.7 Regional EMAP (R-EMAP)

Four interviews were conducted with Regional EMAP (R-EMAP) researchers. A short session at a national meeting of R-EMAP coordinators from the EPA Regions provides an overview; and individual interviews with three Regions were conducted to obtain more detailed information.

B.7.1 R-EMAP National Coordinator's Meeting

November 6–7, 1997

Participants

U.S. EPA ORD R-EMAP Coordinator

- Anthony Carlson, Duluth (218)720-5523 carlson.anthony@epamail.epa.gov

U.S. EPA Regional R-EMAP Coordinators

- Ray Thompson, Region I (617) 860-4372 thompson.ray@epamail.epa.gov
- Darvene Adams, Region II (732) 321-6700 adams.darvene@epamail.epa.gov
- Rollie Hemmett, Region II (908) 321-6756 hemmett.roland@epamail.epa.gov
- Thomas DeMoss, Region III (410) 573-2739 demoss.thomas@epamail.epa.gov
- Jerry Stober, Region IV (706) 355-8705 stober.jerry@epamail.epa.gov
- Arthur Lubin, Region V (312) 886-6226 lubin.arthur@epamail.epa.gov
- Charlie Howell, Region VI (214) 665-8354 howell.charlie@epamail.epa.gov
- Lyle Cowles, Region VII (913) 551-5042 cowles.lyle@epamail.epa.gov
- Jill Minter, Region VIII (303) 312-6084 minter.jill@epamail.epa.gov
- Bob Hall, Region IX (415) 744-1936 hall.robertk@epamail.epa.gov
- Region X (not represented)

ORD

- Barbara Levinson, NCERQA (202) 260-5983 levinson.barbara@epamail.epa.gov
- Rick Linthurst, RTP (919) 541-4909 ricklinthurst@epamail.epa.gov
- Tony Olsen, Corvallis (541) 754-4790 tolsen@mail.cor.epa.gov
- Steve Paulsen, Corvallis (541) 754-4428 paulsen@mail.cor.epa.gov
- Kevin Summers, Gulf Breeze (850) 934-9244 summers.kevin@epamail.epa.gov

EMAP–IM (AED)

- Stephen Hale, Narragansett (401) 782-3048 hale.stephen@epamail.epa.gov

Technology Planning and Management Corporation, ITAS Contractor

- Jeff Rosen (781) 544-3085 jrosen@tpmcscituate.com
- Dillon Scott (781) 544-1298 dscott@tpmcscituate.com

Purpose
The purpose of this meeting was for regional R-EMAP coordinators to meet with ORD EMAP program directors. At the meeting, EMAP–IM (AED) gave a briefing on the status of EMAP information management (IM) and asked for feedback on how information management can be conducted for R-EMAP projects.

Status of Regional Studies
Each regional representative provided an update on the status of their projects and the data produced.

Region I
Please see Section B.7.2, R-EMAP Region I, for complete summary.

Region II
Round 1 was a baseline study of New York/New Jersey Harbor sediments, and has been completed except for the report. The data from this project were managed by Versar and have now been submitted to EMAP–IM (AED). They will be moved to the EMAP Public Web Site as soon as the final report is finished (the report is needed to understand the data). Only the original QA'd data of all measured constituents (e.g., PCB congeners) will be posted. Data analyses and aggregates will not be posted because methodology can change and the results might not be considered valid in the long term.

Round 2 continues the work in the Harbor in a trend assessment of sediment quality and development of indicators. This work will be fully implemented in the summer of 1998 and is the basis for the New York/New Jersey Harbor Estuary Program's long term monitoring program for the Harbor.

Region III
Region III studies focus on the state of streams in the Mid-Atlantic Highlands, including all geographic areas of the Region except the coastal plain and the Piedmont. Data have been produced for biology, habitat, and water chemistry in watershed units. Stressors have been ranked by their contribution to biological indicators. The results were presented to the Regional Administrator and the Assistant Administrator for Water (Bob Perciasepe), who liked the watershed approach. This program is now supported by the Congressmen from Pennsylvania and West Virginia in the Highlands Action Program. The Region is also interested in learning to use landscape ecology methods through technology transfer to results from the Landscape Ecology group. The data from this project has been submitted to EMAP–IM for posting on the MAIA Web Site (under construction) (MAIA 1998).

Region IV
The main effort is the Everglades Ecosystem Assessment. This is a system-wide (4000-mi^2) research and monitoring study of the Everglades ecosystem which has collected data (1993–96) relating to

mercury contamination, eutrophication, habitat alteration, and hydropattern modification issues affecting the system. This study was initially begun to address the mercury issue. However, since the project has developed a comprehensive database, it strongly supports the federal and state Everglades restoration efforts and will provide a means to evaluate present and future management actions related to the much larger restoration efforts. Planners estimate that over the next several years as much as $12 billion will be spent to restore this system. A statistical survey design was used to select 200 canal and 500 marsh sampling stations, a quarter of which were sampled during successive wet and dry seasons over two years. Approximately 30 parameters have been measured in water, soil/sediment, and biological tissue which have been under analysis since sampling terminated. A final technical report is planned for June 1998. Steve Rathbun (University of Georgia) is assisting with statistical analysis.

Region IV has organized these data into ten spreadsheet data files which are undergoing final QA/QC assessment and final proofing and verification (should be completed in January 1998). They have had several requests for the database from other agencies in South Florida; however, they have declined until the metadata can be added to support the files. The goal is to place the data on the EMAP Public Web Site. They need guidance on what is required to get this done recognizing that they have already done a tremendous amount of work on the database. The Region has minimal resources to expend on database management, and they need effective EMAP guidance to allow them to quickly complete the work so it can be released.

The Savannah River project database will probably be handled the same as that for the Everglades. It is probably in QuattroPro and should be posted on the EMAP Public Web Site.

Region V

Art Lubin reported that the Round 1 Corn Belt project is finished, and the data are now available.

Round 2 projects are just getting started because of bureaucratic problems with setting up contracts and cooperative agreements. The projects will develop biological indicators for watersheds and assess the status of wadeable streams in order to understand the spatial evolution of northern lakes and forests. They will be comparing random sites to intensive sites. Sampling will begin in Spring 1998.

No data submission standards were included in the contracts, and the project managers at Region V will be raising data submission issues with the principal investigators at a meeting later this year.

Region VI

Studies in 1993 and 1994 focused on Toxic Substances Characterizations for Selected Texas Estuaries, which is a follow-up to the EMAP–Estuaries (Louisianan Province) monitoring program. The 1993 project was in Galveston Bay, the 1994 project in Corpus Christi Bay. Both projects collected the full suite of EMAP indicators (PCBs, metals, PAHs, TBT, etc.). Data management for

the projects is being handled by Gulf Breeze in a manner similar to the EMAP Estuaries data. The Toxic Substances studies were designed to:

- estimate the degree, extent, and potential effects of tributyl tin concentrations in sediments of Galveston Bay and relate them to the condition of the benthic macroinvertebrate (bottom-dwelling organism) communities;
- estimate the type and rate of fish pathologies (eye deformities, skin lesions, organ abnormalities, etc.) found in the East Bay Bayou of Galveston Bay and relate these pathology findings to estimated sediment toxicity and contaminant levels;
- estimate the levels, extent, and distribution of contaminants in fish and sediments in tidal reaches of the Arroyo Colorado and the Rio Grande; and
- estimate the levels, extent, and distribution of contaminants in fish tissue and sediments in Corpus Christi Bay.

The Galveston report has been distributed for peer review. The Corpus Christi work is still in progress; field work has been completed.

The 1996–1997 study is the application of a probabilistic approach to determine the extent and effects of stream habitat degradation and fish community integrity in eastern Texas streams. The project is now being designed in cooperation with ORD-Western Ecology Division. The project is collecting data on biological indicators, physical habitat, and water quality. This study will use EMAP methodology to:

- evaluate the status of physical habitat and biological communities in small, perennial east Texas streams;
- determine the extent of degradation by land use and among ecoregions;
- further refine biological criteria and habitat evaluation techniques for this region; and
- determine the prevalence of mercury contamination in fish tissue from small streams.

The Texas Parks and Wildlife Department River Studies Group will conduct the study. Field work will be conducted during 1997–1999, and a final report is expected during the year 2000.

Region VI will input the data into STORET so that it may be readily aggregated with and segregated from data sets generated through other programs (fields will be available in STORET to separate "EMAP" and "R-EMAP" data from the other stuff).

Summaries of the two projects are available on the web (REMAP 1998a, REMAP 1998b).

Region VII
Lyle Cowles reported that Round 1 focuses on the status of stream water quality of Nebraska, Kansas, and Missouri. This project is developing an IBI for fish and habitat to give the states a measurement tool for the health of fisheries in fresh water. They sampled water, fish tissue, and sediment (not macroinvertebrates) for chemistry. They are now analyzing the data and will have a draft report by March 1998. EMAP has already collected the data at Corvallis.

The Round 2 study will add a land use/land cover component to the first project for a landscape analysis and characterization to support regional environmental assessment. They will be collaborating with the Kansas Center for Applied Remote Sensing (KCARS) on this work.

The most recent project is resampling Nebraska streams from the Round 1 study (approximately 60 randomly selected) and assisting the state of Nebraska with using probabilistic design and EMAP indicators for rotating basins studies. Region VII will also be assisting the states with analyzing the data and writing reports.

There is a variety of monitoring work being done in most of the states in Region VII. However, because the monitoring is not coordinated spatially, temporally, or for quality or quantity, it is not possible to use it for estimating the overall health of resources.

Region VIII
See Section B.7.3, R-EMAP Region VIII, for a complete summary.

Region IX
See Section B.7.4, R-EMAP Region IX, for a complete summary.

Region X
No representative attended.

R-EMAP Information Management Issues
The EMAP Director stated that it is important for R-EMAP projects to make their data available through the EMAP Public Web Site. After the Regional Reports, EMAP–IM (AED) gave an overview of data management issues and asked for feedback from the Regions, and indicated that there will be follow-up by phone after the meeting.

Since the money earmarked for data management in Round 2 was never delivered to ORD and the Regions, participants indicated that EMAP–IM (AED) can assist with documentation and distribution (EMAP-IM is looking for funds to add staff to assist with this work). EMAP–IM recommended that future budgets would need to set aside 10–20% in order to handle data management, documentation, and distribution tasks.

The ITAS Contractor (TPMC) reviewed the data management issues that EMAP needs to understand to revise the IM Plan, including such issues as where R-EMAP data will reside and what assistance the projects need from EMAP–IM to make their data accessible. The issues of importance include:

- **Data management**—how data are checked for quality and completeness;
- **Data products**—what data sets and other products of collection and analysis are being produced;
- **Data distribution**—how data sets will be made available beyond the data collectors;
- **Data Documentation**—how the quality and details of data sets can be described for future users;
- **Project management**—how does project structure affect the success of information management and access; and
- **How EMAP–IM can help**—what resources, expertise, and guidance can be provided to help complete R-EMAP information management.

Important points made in this discussion included:

1. The EMAP Director pointed out that for R-EMAP data to be valuable it must be well-documented and available to researchers beyond those who collected the data. EMAP–IM provides an option for archiving, documenting, and distributing data that researchers should use. This effort is especially important for monitoring data sets.

2. One of the most important features of the EMAP–IM system should be simplicity—the ability of the user to get to desired data easily through simple interfaces without understanding the underlying complexity of the data. It is also very important for EMAP–IM to simplify the data management and access process for the researchers.

3. Round 1 projects had no contractual requirements for submitting data sets, so EMAP–IM can assist the Regions with collecting them. Round 2 had some guidelines that were given to the R-EMAP coordinators; Round 3 could benefit from additional planning and resources.

4. R-EMAP must do effective data management and distribution in order to be credible as a national program. R-EMAP coordinators could benefit from guidance on when and how data documentation should be done, especially if the project manager changes during the period of the project.

5. Creating Data Directory entries for every R-EMAP data set should be a minimum requirement (they only take about 10 minutes to complete).

6. Metadata creation tools are becoming more available and improved for creating FGDC-compliant data. R-EMAP groups can get assistance from EMAP–IM and others who are familiar with these tools and standards.

7. The Regions do not all have sufficient hardware, software, infrastructure, expertise, and resources to manage and distribute R-EMAP data. Some will distribute data through partnerships with research partners and data repositories, and others will need assistance from EMAP–IM.

8. It would be very helpful if EMAP–IM could provide data submission guidelines for subcontractor agreements to improve the delivery of data sets.

9. A bibliography of EMAP publications and online access to electronic publications would be extremely useful (the EMAP Director indicated that EMAP is now creating a searchable bibliography of its publications).

10. Participants can let EMAP–IM know about non-EMAP data sets that should be tracked in the Data Directory (i.e., a PI has a data set available on a university web site).

11. R-EMAP projects would benefit from EMAP efforts to index methods, documents, and final reports from other Working Groups.

Conclusions and Next Steps

Although data management is stated to be an important goal of R-EMAP, additional resources are required for coordinating data submission, management, dissemination, and analysis. It is important to plan adequate resources to ensure that data are made available with high-quality documentation so that the results can be aggregated for regional analysis. Each Region handles data management differently. Although they are planning to submit data to EMAP–IM, they need guidance and assistance from EMAP–IM on data submission, management, documentation, creation of Data Directory entries, and distribution.

EMAP–IM will assist each region with delivering data to be posted on the EMAP Public Web Site.

The R-EMAP portion of the EMAP Public Web Site needs to be brought up to date with Round 2 projects. Ron Carlson will supply EMAP–IM with this information.

It will be important for EMAP–IM to set up links to national and regional databases where R-EMAP projects will be storing their data. Most data on water, sediments, and biology will be entered into modernized STORET as soon as it becomes available. Diane Switzer of Region I is well-informed about the new STORET and is a resource for other groups on this topic. Air monitoring data from Region I will be entered into a number of national air monitoring databases. Participants support the

idea of entering EMAP data into STORET because EPA has spent millions of dollars to update this system and efforts are underway to accommodate biological and habitat data that are consistent with EMAP methodologies. With the recent demise of NWIS II, USGS will likely adopt STORET as their primary database. Fields will be available to separate EMAP and R-EMAP data from the other program's data. Region VI will input their future data into STORET so that it can be readily aggregated with and segregated from data sets of other programs.

EMAP–IM and the ITAS contractor followed up this meeting with phone interviews of selected regional staff to learn more about the status of data and information management in the projects for summaries (see Sections B.7.2, B.7.3, B.7.4).

B.7.2 R-EMAP Region I

Summary of Workshop
September 16, 1997

Participants

U.S. EPA Region I Laboratory, Lexington, MA

- Diane Switzer
- Ray Thompson
- Alan Van Arsdale

Technology Planning and Management Corporation, ITAS Contractor

- Jeff Rosen
- Dillon Scott

Mission
The focus of these studies is the investigation of mercury and other toxics in fish, water column, sediments, and atmospheric deposition in New England lakes and streams.

Data Collection Efforts
There are currently three projects in two rounds of funding.

Round 1 (1993–1994)
Fish Tissue Contamination in the State of Maine (Maine Department of Environmental Protection, Maine Department of Inland Fisheries and Wildlife, U.S. EPA Region I Environmental Services Division)—

This project was designed to evaluate the distribution of mercury in surface waters, fish, and sediments. All sampling was performed on the EMAP sample frame (developed by Tony Olson), which originated with 1,800 lakes. Of the lakes in the original frame,150 candidate lakes were chosen with a goal of sampling 120–125 lakes. A total of 125 lakes were sampled over the 2 years.

Most of the work was performed by the Maine Department of Environmental Protection (technical contact Barry Mower) and the Maine Department of Inland Fisheries and Wildlife.

Fish sampling was performed using regular long-line fishing with gill nets. The objective of the fish collection was not to determine populations but to collect tissue specimens. Fish tissue analyses for mercury were done on whole fish and filets to determine differences between them. For the sediment analysis, some subset of the EMAP suite was analyzed, including lead, cadmium, and mercury.

A major outcome of the study was a Maine advisory on mercury in freshwater fish. The findings inspired other New England states to study mercury in their own lakes and resulted in policy changes and regulations at the state and regional levels. Each state used a different methodology based on available funds (e.g., New Hampshire asked fishermen for samples, Massachusetts DEP and DPH collected samples). The data were also used in regional (New England Interstate Water Pollution Control Commission, Northeast States for Coordinated Air Use Management, Northeast Waste Management Officials Association) and national (Office of Air Quality Planning) assessments, and some link was made also with Canadian data. The work also contributed to some spin-offs including the Casco Bay Program, the Regional Applied Research Effort (UNH, ORD), and others.

Data Types

General data types are listed below. Please see the **Additional Information On Data Types** subsection, below, for more detail.

- Lake Field—Sampling Data
- Fish Collection Descriptive Data
- Water Quality Profiles
  - ❑ QA/QC data from samples run by the labs (U.S. EPA Region I, New England Regional Laboratory; Maine Health and Environmental Testing Laboratory for Inorganic and Organic Parameters; National Biological Survey U. of Maine Laboratory; U. of Maine Sawyer Environmental Chemistry Laboratory)
- Organic compounds measured include:
  - ❑ Aldrin, A-BHC, B-BHC, D-BHC, G-BHC, A-Chlordane, G-Chlordane, Dieldrin, Endosulfan I, Endosulfan II, Endosulfan Sulfate, Endrin, Endrin Aldehyde, Endrin Ketone, Heptachlor, Heptachlor Epoxide, DDE, DDT, DDD, Toxaphene, Aroclor 1221, Aroclor 1232, Aroclor 1242, Aroclor 1248, Aroclor 1254, Aroclor 1260, Aroclor 1268, Percent Surrogate Recovery, Percent Moisture, Percent Lipids

Round 2 (1997–)

Measurement of Mercury Deposition and Atmospheric Concentrations in New England (with Northeast States for Coordinated Air Use Management), and Assessment of Mercury in

Hypolimnetic Lake-Bed Sediments of Vermont and New Hampshire (with Vermont Agency of Natural Resources and New Hampshire Department of Environmental Services)—

The two Round 2 R-EMAP projects are off-shoots of the Round 1 project and will consist of two projects now being initiated. A major objective of Round 2 is to allow testing of multiple hypothesis regarding the air quality and the effects of atmospheric deposition on a number of species in surface water systems. Attempts will be made to correlate atmospheric deposition with sources for emissions. It is anticipated that the data analysis for Round 2 will be performed by a consortium of principal investigators using ancillary funds, and that the data will be managed via the Internet and workshops. The workshop approach will be to distribute the data with the questions that are being asked, and then request that principal investigators perform analyses to present to their peers at the workshop.

Round 2 will have intensive sites monitoring for ozone precipitation as well as concentration of some species of chemicals in the precipitated water. There is no expectation that these intensive sites will in any way be coordinated with the EMAP Intensive Sites. In addition, Round 2 will include measurements far beyond those made in Round 1 and will not only include chemical species but also particle analysis for two components greater than 2.5 microns and less than 2.5 microns. In addition, cloud monitoring will be performed on Mount Mansfield. No standards are in place for this since methodologies do not currently exist for analyzing mercury in clouds.

Data Types
Only general information is available because these projects are just beginning.

NESCAUM project—

- Anticipated mercury data to be collected in the New England region include:
  - ❑ Approximately 16 sampling locations for mercury monitoring (wet, gaseous, particle) network
  - ❑ Potential deposition gradients
  - ❑ Spatial and temporal patterns of deposition
  - ❑ Atmospheric concentration
  - ❑ Local vs. regional influences on atmospheric concentrations
  - ❑ New method for measuring mercury in mountain clouds
  - ❑ Data from new method on whether mountain-cloud impaction results in enhanced mercury deposition at high elevations
  - ❑ Event-specific data on mercury precipitation that may be used to evaluate and develop atmospheric transport and deposition models or that may be used in conjunction with trajectory models, source-receptor models, and other pollutant

data to assess significance of individual source categories or source regions to deposition in New England

New Hampshire/Vermont project—

- Location and descriptive data for 90 Vermont and New Hampshire lakes
- Total and methylmercury concentrations in water and surficial sediments in 90 lakes
- Fish-tissue mercury levels in 20 of study lakes
- Relationship between aqueous and sediment total and methylmercury concentrations, fish-tissue total mercury concentrations, and physico-chemical lake and watershed characteristics
- Concentrations or conditions of water chemistry parameters that mediate mercury methylation in 90 lakes
- Evaluate (rank) potential risk of migration of total sediment mercury into water column in methylated form
- Investigate historical deposition patterns of total mercury in dated cores from six lakes; relate these depositional patterns to known historical events in surrounding watersheds where possible. Also investigate deposition profile of total mercury in undated cores from the sediments of 12 lakes (in addition to the 6 lakes above). Compare stratigraphy of mercury from Vermont and New Hampshire lakes with that of selected Maine and Adirondack studies

Users

R-EMAP data users include:

- Primary users:
  - ❑ R-EMAP researchers;
- Secondary users:
  - ❑ Researchers and planners doing syntheses at regional, national, international level; and
- Data repositories:
  - ❑ Region I Laboratory
  - ❑ Mercury Deposition Network
  - ❑ University of Michigan
  - ❑ NEARDAT (George Washington University).

Water research data will be used at the state level and in other research projects on those water bodies. The water, fish, and sediment data have so far been used only by the researchers, but it is anticipated that the data will also be used by:

- Regional offices and national R-EMAP program for regional and national assessments
- State, regional, local planners
- Watershed groups
- Legislators, policy makers

Air monitoring users include a network of collaborators who archive and use the final data and methodology into regional information networks.

Data Sources

Round 2 will need to use information from a number of different source databases and inventories, including local data, EMAP lakes and streams data, EMAP indices, and other EMAP approaches and methodologies. Data will also be retrieved from the EPA Aerometric Information Retrieval System (AIRS), and further data will be sought for air deposition and from facilities (potential point sources).

Interviewees emphasized the need for EMAP–IM to provide access to summaries of approaches, equations and methodologies for developing indices—particularly the Habitat Quality Index developed for the mid-Atlantic. The development of assessment tools has always been a strength of EMAP and should be available to the Regions and States.

Data Management

Data are quality-assured, documented, analyzed, and maintained by the state agencies and regional boards conducting the projects. The quality and longevity of stewardship for these databases at the project sites is unknown and will vary by site. Round 1 data reside at MEDEP. Region I's understanding of the original data sets is that there is a single record for each fish sampled. Data formats at the project site (e.g., spreadsheets, GIS coverages) are unknown. The Maine project was required to submit data to EPA under verbal guidance from Region I through existing QAPP requirements. Round 2 data management will be conducted by the states and NESCAUM, who will conduct the research. Data management requirements have been discussed, but no firm plans have been put into place. Analytical protocols have been standardized, and it is expected that this standardization will also result in a standard for data submission. The participants will adopt the Mercury Deposition Network (MDN) format as the standard data format. Round 2 has an international component, since some of the work will be done around the Great Lakes, and there will be a need to exchange data with Canadian sources. This exchange presents problems with both

methodology as well as data management since the protocols between the U.S. and Canada are not standardized.

QA/QC'd raw data will be submitted by the project managers to Region I in ASCII format. There have been no problems obtaining the data from Round 1 because of good long-term working relationships between Region I and the projects (no problems are anticipated for Round 2). However, Region I R-EMAP has sufficient resources allocated for data collection and analysis but very little for data management and distribution. The projects will produce useful data aggregates and technical reports that should be stored and disseminated. In addition, the water data should be entered into STORET. However, the Region currently only has sufficient resources for storing the ASCII files.

Data Quality Assurance/Quality Control
All field QA/QC is done at the project sites, and procedures are documented in the reports and QAPPs.

Data Products
Round 1 data products delivered to Region I consist of tables in technical reports and ASCII data files.

Round 2 data products delivered to Region I will include tables in assessment reports and ASCII data files.

Data Formats
Region I maintains the ASCII files on disk. Data formats at the project sites are unknown.

Data Distribution
It is anticipated that the water, fish, and sediment data from both Round I and Round 2 will ultimately reside in the new STORET (STORET X) system. Region I has evaluated a prototype of the data entry system for the STORET X and found it to be an improvement over the earlier version. Many of the issues regarding the loading of the data and production of reports have been eliminated, and the new system appears to be user-friendly and capable of multiple analyses.

Air data will be stored and distributed in a number of different ways, some of which are still being worked out among the researchers and the data networks in which they participate (an upcoming meeting will address this issue). Some of the atmospheric mercury data will be loaded into the EPA AIRS database. Other data which are too detailed or require extensive precision and accuracy will not fit well into AIRS, and these data are anticipated to be stored by many of the principal investigators. Some of the air/precipitation data and metadata from Round 2 will also make its way into the MDN and University of Michigan sites. Other atmospheric data will be transferred from the project sites to the NEARDAT web site at George Washington University (contact is Rudy Husar). [George Washington University is also the repository for the Ozone Transport Assessment Group

(OTAG) data.] These data will be scattered among a number of air database repositories and will have to be carefully tracked in the EMAP Data Directory so that users will know where to find the data.

The lack of available resources for entering the data into sites such as STORET is also a concern. The Region I project managers indicated that it would be helpful if EMAP–IM could make the R-EMAP data files temporarily accessible on the EMAP Public Web Site until they can be entered into the appropriate repositories. In the future, if the data are archived in other repositories, the EMAP Data Directory could be linked to them.

Documentation
The cooperative agreements in Round I did not require projects to prepare metadata, so metadata will have to be prepared by Region I from project materials (QAPPs, notebooks, databases) when data are entered into repositories (STORET, AIRS, etc.). None of the quality assurance or documentation of methodologies for gathering and analyzing samples is in a form amenable for electronic storage and access. This documentation will need to be reformatted so that it is accessible via STORET or the EMAP–IM.

Software
Data analysis takes place at the project sites. No special software is needed at Region I to store the ASCII files.

Hardware and Infrastructure
Region I has sufficient Internet access and computing resources to store the ASCII files.

Project Management
Research and data management are mainly conducted by the project sites, who collect, analyze, use, and store the data. Region I receives final reports and data and will load it into appropriate repositories (EMAP–IM, STORET, AIRS, etc.) in the future.

The Region I project managers indicated a concern regarding program management. Currently, with the way the money for R-EMAP is routed, the money cannot actually come to the Regions, although the Regions are ultimately responsible for coordinating the projects. The actual project officers for all of these projects are in ORD where the money originates. The theoretical flow of authority is from the coordinator to the project manager to the cooperators. The real path is from the coordinators to the cooperators with the project manager in a money management and contract management role but not in any technical role. This is a result of the inability of the Regions to obligate funds originating within ORD.

How Can EMAP–IM Help?

The participants were asked to evaluate the planned EMAP–IM approach for its potential usefulness to them, and to indicate what other services EMAP–IM could provide. They indicated that it would be a major contribution if EMAP's Data Directory can cross-reference and index data sources and then facilitate the access to the data sources. They also indicated that posting the R-EMAP data on the EMAP Public Web Site would be a great service if it included long-term maintenance. Long-term maintenance would include assuring that the data were properly stored and available to interested parties. They indicated that the Surf Your Watershed information system currently available on the EPA Internet site would also be useful for making R-EMAP data available.

EMAP–IM could help Region I R-EMAP by providing them with data aggregates and methodology from other EMAP efforts, especially the assessment tools (e.g., Habitat Quality Index, IBI, indices, actual equations, other tools) discussed in the **Data Sources** subsection. These assessment tools are currently difficult to locate and obtain.

EMAP–IM can also help by providing links to existing data networks (e.g., MDN, George Washington University), and by indexing and making available those data sets (e.g., some of the air data) that are retained only by principal investigators and not incorporated into other networks. They would also be interested in learning more about other programs like the Great Waters program (which include the Great Lakes and Lake Champlain).

A general observation is that STORET is likely to play a key role in much of the water, tissue, and sediment data generated from both of these projects. The point was made that the national core performance program requires that states generating selected data upload these data to STORET. The Region I project managers indicated that for New England, there is currently no central information database or server where information can be placed for access by other groups.

Other services that would be useful for EMAP–IM to supply would be sets of standard forms, formats for proposals, and tools for generating appropriate documentation.

The participants also are concerned that the institution of themes in Round 3 may inhibit projects within individual regions. R-EMAP funds are the only source for conducting local and regional monitoring and research programs. The purpose of R-EMAP is to address regional problems, so each Region must be somewhat different. Approaches that address the needs of one Region will not necessarily be useful to another Region. For example, Region I projects are organized around common regional concern with mercury deposition.

A final point made by the participants is that data management for R-EMAP projects is currently not coordinated. It is stated to be an important EMAP goal, but there are few resources for coordinating data submission, management, dissemination and analysis.

Conclusions and Recommendations

An overall observation regarding this interview was that the complexity of EMAP programs does not seem to change with scale. Data management, particularly for Round 2 of this R-EMAP program—which includes multiple parameters, media, and research collaborators—is easily as complicated as data management issues for the overall EMAP program. It would be a mistake to assume that the requirements for information management in working groups or partner programs is of a lesser degree of complexity and therefore less expensive to implement than those in larger regional/national programs. Ignoring these data management issues will result in R-EMAP data sets being dispersed with little coordination, and will result in an inability to aggregate these results within EMAP. At the very least, R-EMAP data sets should all be indexed at a single site and cross-referenced so related data sets can be found. In addition, data dictionaries and recommended standards should be published with the data sets so that there is a better chance of integrating this information into a larger context.

Additional Information On Data Types

Round 1

Maine—

- Lake Field Sampling Data
  - ❑ Lake descriptive information (locations, names, morphometric data)
  - ❑ Cation, Anion, Air Equilibrated pH, Acid Neutralizing Capacity and True Color Data
  - ❑ Total Phosphorus
  - ❑ Dissolved Organic Carbon
  - ❑ Sediment Analysis Results
  - ❑ Sediment Percent Solids—Comparison of Split Sample Results
- Fish Collection Descriptive Data
  - ❑ Results of Analyses for Organic Compounds, Percent Moisture, and Percent Lipids in Whole Fish
  - ❑ Mercury, Cadmium, and Lead Concentrations in Fish Tissue Composites
  - ❑ Summary of Organic Compounds Detected in Whole Fish Composites
  - ❑ Split Sample Results (Summary of Split Sample Results; Mercury and Percent Moisture in Predator Fillets—Split Sample Results; Metals in Whole Fish—Comparison of Split Sample Results; Organic Compounds, Percent Moisture, and Percent Lipids in Whole Fish—Comparison of Split Sample Results; Metals in Sediment—Comparison of Split Sample Results
- Water Quality Profiles

- ❑ Field Duplicate Results (Summary of Field Duplicate Sample Results; Water Quality Profile Duplicate Results; Anion and Cation Field Duplicate Results; Total Phosphorus and Dissolved Organic Carbon Field Duplicate Results; Sediment Field QA/QC Results

- U.S. EPA Region I, New England Regional Laboratory QA/QC Results
  - ❑ Summary of reporting limits, summary of metals analyses QA/QC, frequency of QA/QC sample analysis for organic compounds, EPA data quality objectives, and summary of QA/QC results for organic compounds
  - ❑ Inorganic Duplicate Samples
  - ❑ Organic Duplicate Samples
  - ❑ Inorganic Spike Samples
  - ❑ Organic Spike Samples
  - ❑ Inorganic Reference Materials (Replicates and Percent Recoveries)
  - ❑ Organic Reference Materials (Percent Recoveries)
  - ❑ Metals in Whole Fish (EPA Splits)
  - ❑ Mercury in Predator Fillets (EPA Splits)
  - ❑ Metals in Sediment (EPA Splits)
  - ❑ Organic Compounds (EPA Splits)
- Maine Health and Environmental Testing Laboratory (HETL) QA/QC Data for Inorganic Parameters
  - ❑ Data quality objectives for inorganic compounds, reporting limits for inorganic compounds, frequency of inorganic QA/QC samples analyzed, and summary of inorganic QA/QC sample results
  - ❑ Mercury in Tissue
  - ❑ Cadmium in Tissue
  - ❑ Lead in Tissue
  - ❑ Laboratory Blanks—Metals in Tissue
  - ❑ Reference Samples in Water Matrix—Cadmium and Lead
  - ❑ Sediment Grain Size
  - ❑ Mercury in Sediment
  - ❑ Lead in Sediment
  - ❑ Cadmium in Sediment
  - ❑ Dissolved Organic Carbon
  - ❑ Total Phosphorus
  - ❑ Equipment Blanks

- Maine Health and Environmental Testing Laboratory (HETL) QA/QC Data For Organic Parameters
  - ❑ Frequency of organic QA/QC samples analyzed, reporting limits for organic compounds, fish composite samples not analyzed for the 50% organic fraction, data quality objectives and summary of organics results, duplicate sample results for percent moisture and percent lipids, organic compound results for standard reference material, surrogate recoveries for organic compounds in fish
  - ❑ Sample results
  - ❑ Duplicate results
  - ❑ Spike added
  - ❑ Spike conc. results
  - ❑ Blank conc.
  - ❑ % rec.
  - ❑ Dup. % Diff
- National Biological Survey U. of Maine Laboratory QA/QC Data
  - ❑ Summary of QA/QC results for mercury and percent moisture in predator fillets
  - ❑ Standard Reference Material
  - ❑ Spiked Samples
  - ❑ Mercury Duplicate Samples
  - ❑ Percent Moisture Duplicate Samples
- U. of Maine Sawyer Environmental Chemistry Laboratory QA/QC Data
  - ❑ Anion and cation target holding times, method detection limits, and precision and accuracy objectives, summary of QA/QC results for anions and cations
  - ❑ Cation Precision Calculations—1993 Samples
  - ❑ Anion Precision Calculations—1993 Samples
  - ❑ Laboratory Blanks—1993 Samples
  - ❑ Laboratory Blanks—1994 Samples
  - ❑ Cation QC Checks—1993 Samples
  - ❑ Cation QC Checks—1994 Samples
  - ❑ Anion QC Checks—1993 Samples
  - ❑ Anion QC Checks—1994 Samples
- Organic compounds include:
  - ❑ Aldrin, A-BHC, B-BHC, D-BHC, G-BHC, A-Chlordane, G-Chlordane, Dieldrin, Endosulfan I, Endosulfan II, Endosulfan Sulfate, Endrin, Endrin Aldehyde, Endrin

Ketone, Heptachlor, Heptachlor Epoxide, DDE, DDT, DDD, Toxaphene, Aroclor 1221, Aroclor 1232, Aroclor 1242, Aroclor 1248, Aroclor 1254, Aroclor 1260, Aroclor 1268, Percent Surrogate Recovery, Percent Moisture, Percent Lipids

Round 2

NESCAUM project—

- Anticipated mercury data to be collected in the New England region include:
  - ❑ Approximately 16 sampling locations for mercury monitoring (wet, gaseous, particle) network
  - ❑ Potential deposition gradients
  - ❑ Spatial and temporal patterns of deposition
  - ❑ Atmospheric concentration
  - ❑ Local vs. regional influences on atmospheric concentrations
  - ❑ New method for measuring mercury in mountain clouds
  - ❑ Data from new method on whether mountain-cloud impaction results in enhanced mercury deposition at high elevations
  - ❑ Event-specific data on mercury precipitation that may be used to evaluate and develop atmospheric transport and deposition models or that may be used in conjunction with trajectory models, source-receptor models, and other pollutant data to assess significance of individual source categories or source regions to deposition in New England

New Hampshire/Vermont project—

- Location and descriptive data for 90 Vermont and New Hampshire lakes
- Total and methylmercury concentrations in water and surficial sediments in 90 lakes
- Fish-tissue mercury levels in 20 of study lakes
- Relationship between aqueous and sediment total and methylmercury concentrations, fish-tissue total mercury concentrations, and physico-chemical lake and watershed characteristics
- Concentrations or conditions of water chemistry parameters that mediate mercury methylation in 90 lakes
- Evaluate (rank) potential risk of migration of total sediment mercury into water column in methylated form
- Investigate historical deposition patterns of total mercury in dated cores from six lakes; relate these depositional patterns to known historical events in surrounding watersheds

where possible. Also investigate deposition profile of total mercury in undated cores from the sediments of 12 lakes (in addition to the 6 lakes above). Compare stratigraphy of mercury from Vermont and New Hampshire lakes with that of selected Maine and Adirondack studies

B.7.3 R-EMAP Region VIII

Conference Call Summary
Friday, December 19, 1997

Participants

- Jill Mintner, Region VIII Coordinator
- Dillon Scott, Technology Planning and Management Corporation

Purpose
To understand R-EMAP data management for Region VIII.

Studies
The studies are behind schedule because the R-EMAP coordinator position was vacant for six months in 1997. The new coordinator is getting the projects back on track and addressing the data documentation and distribution needs.

The Round 1 project is "Assessment of Metals Impacts in Headwaters Streams within Mineralized Areas of the Southern Rockies Ecoregion," a study of total metals release from abandoned mines. Its purpose is to develop a picture of the extent of the impacts, environmental indicators, and reference conditions for mineralized streams. Sampling was completed on an EMAP design. Colorado State University (CSU) has now analyzed the macroinvertebrate data and worked up a Regional Biotic Index to assess the effects of heavy metals (this work was presented at the November 1997 Society of Environmental Toxicologists and Chemists meeting). ORD-Western Ecology Division collected and is analyzing data on water chemistry, fish, and sediment data. The R-EMAP coordinator will prepare a final report on the macroinvertebrate data and will document and archive the data. In the future, the Region will try to analyze the ORD data, put it into a report, and document it if additional information and resources become available.

The Round 2 project is now in the early stages of planning. The potential project will focus on grazing impacts on rangeland conditions in Utah. It is being managed by Roger Dean, who is in the nonpoint source group at Region VIII. It has not yet been funded, but will be conducted by Utah State University and ORD Characterization Research Division (Las Vegas, Dan Heggem), with some

involvement by the Bureau of Land Management. The project does not conform to EMAP sampling design and is still under review.

For Round 3, Region VIII would like to accomplish the following goals:

- use a portion of the funds to complete data management and metadata for Round 1 (including producing the Final Report, and archiving and documenting the data); and
- get the States actively involved in developing monitoring design and data analysis for a project so they are on board with sampling design, analysis, and assessment methods; evaluate possible application of the methods to State programs.

Users

The primary users of Region VIII R-EMAP data and Regional Biotic Index for macroinvertebrates are:

- EMAP researchers at Universities and ORD;
- States; and
- EPA scientists and managers.

Data Sources

Southern Rockies: Most of the data were newly collected for this project. However, the Regional Biotic Index was created using data from previous studies conducted by the principal investigator at CSU (Will Clements).

Utah Rangelands: Region VIII has requested land cover data that will not be ready until the year 2000. The project will create a sediment transport model and collecting water chemistry data.

Region VIII will also need access to the following data for future work:

- characterization data for the Region, since the State monitoring programs only sample impaired areas ("targeted" sampling design);
- macroinvertebrate, fish tissue, sediment data to characterize biological integrity, and assess human health risks; and
- MRLC data for the Utah Rangelands project, to link impairment with land use, and for relative ranking of stressors.

Region VIII would like assistance with accessing EMAP 1990–1995 data sets that are relevant to their studies.

Data Management

Southern Rockies: The data are managed by data collectors (CSU and ORD). CSU uses spreadsheets for data management and storage, and enters the data as needed, through ASCII conversion, into SAS for analysis. The data can be provided to the Region in spreadsheet format. ORD stores their data in SAS and converts it to ASCII for the Region. However, when they provided the ASCII files of fish, water chemistry, and sediment data to the Region, the data could no longer be matched up to station numbers, so the Region is trying to fix this problem now. ORD provided information about what EMAP protocol was used for each data type, and will need to supply further information (e.g., laboratory, methods, analyst) and metadata about how the data were collected and processed. ORD will also deliver the habitat data to the Region.

Utah Rangelands: Data will be managed by the data collectors. Region VIII will ensure that data and documentation are submitted by project managers. Region VIII would like guidance from EMAP–IM on data submission standards that could be included in the contract.

Data Quality Assurance/Quality Control

Southern Rockies: Data QA/QC was done by data collectors. The Region will need to examine whether the percentage of impaired waters was calculated appropriately using EMAP statistical methods.

Data Products

Products of the Southern Rockies mineralized streams study include:

- Regional Biotic Index;
- Final Report (in progress);
- Raw data and its metadata (in progress);
- Abstracts from SETAC meeting (Clements); and
- Published papers (Clements).

Data Distribution

Southern Rockies: Region VIII will provide long-term access to the data through the following products:

- paper reports;
- metadata;
- archived data (in STORET and at Region);
- links from EMAP to data archives; and

- EMAP Data Directory entries.

Macroinvertebrate data will be archived at Region VIII and prepared for entry into the modernized STORET system. Data can also be made available to EMAP–IM for inclusion on the EMAP Public Web Site, and a link could be added to STORET when those entries are made. The ORD data will have to be cleaned up, documented, and analyzed before it is distributed.

The R-EMAP coordinator mentioned that it will be especially important to provide data to the States in a database they can access without having to go through the Region. The States have legal authority and responsibility for the resources, so it is important for them to see their data and the data from other states. They are also the group that will share the data with the public. The States also need access to the Regional Biotic Index for assessing metals-impacted streams.

Documentation
Metadata for macroinvertebrate data is being prepared by the R-EMAP coordinator and a Region VIII data systems coordinator. ORD needs to provide metadata for their data types.

EMAP Data Directory entries will have to be prepared by the R-EMAP coordinator or the Region VIII data analyst who is assisting with the ORD data.

Software, Hardware, Infrastructure
This project has been receiving assistance from Region VIII information systems staff on data archiving and documentation, and so far this support has been sufficient because the project does not generate large, complex data types. However, additional resources may be required in the future if landscape ecology data are generated for the Utah Rangelands project.

Project Management
The R-EMAP coordinator will pull together the data, metadata, and Final Report for the Southern Rockies study with assistance from a Region VIII data analyst. The Utah Rangelands project is being coordinated with a colleague in the nonpoint source group at Region VIII. The R-EMAP coordinator will ensure that R-EMAP goals are met and that data and metadata are collected for the project.

How Can EMAP–IM Help?
EMAP–IM can provide data submission standards for the Utah Rangelands study, and assist with incorporating ORD data and metadata for the Southern Rockies study.

B.7.4 R-EMAP Region IX

Conference Call Summary
November 21, 1997

Participants

- Robert Hall, Region IX Coordinator
- Dillon Scott, Technology Planning and Management Corporation

Purpose
To understand R-EMAP data management for Region IX.

Studies
There are currently three projects in two rounds of funding.

Southern California Bight: The first study was an assessment of the Southern California Bight from Pt. Conception to the Mexico border. An EMAP sampling design was used for this project, and almost all of the EMAP estuaries parameters were collected. The EMAP design was also subsequently incorporated into the new International Treatment Plant's pre-discharge monitoring program.

Central Valley, California: The second study is in California's Central Valley, and consists of an EMAP surface water assessment in natural streams and constructed conveyances. The EMAP sampling design was employed here to collect the standard set of EMAP surface water parameters. This is an area of high water demand with water supply controlled mostly by agricultural corporations. Residence time of water is an important determinant of pollution effects; water is generally moved among fields rapidly and spreads contaminants before they can filter out. However, the project was unable to calculate residence time very well because irrigation managers often change the flow in the channels. Data analysis for this project is being completed with help from ORD-Western Ecology Division. Habitat data should be done soon.

Humboldt River, Nevada: The third project is in the Humboldt River watershed, Nevada, and focuses on aquatic systems in the Basin and Range Province where dewatering issues are of concern. This work is being done by Dr. Gary Vinyard, University of Nevada, Reno. The project is now in the design phase and will be developing new EMAP protocols for arid assessments. They are now modifying the EMAP Surface Waters sampling design for an arid habitat, and will be using the Office of Water's Rapid Biological Assessment (Mike Barbour *et al.*).

Region IX is very interested in the contribution that Landscape Ecology can make with landscape characterization (Bill Kepner). They also would like more uplands involvement.

Users
The primary users of Region IX R-EMAP data are:

- EPA Region IX environmental program managers (ag team, NPDES wastewater permits, Border, etc.);
- project researchers at Universities;
- ORD researchers at Corvallis; and
- State of California, so they can pick up the Central Valley program.

Data Sources
Southern California Bight: Pre-existing information included NPDES discharge monitoring data from participating POTWs, as well as historical (1977, 1985, 1990) data from SCCWRP.

Central Valley, California: The only available base data was EPA River Reach (RF3) files. QA on this data source (checking 300 sites) showed that there is a 20% error rate in the coding (e.g., segment coded as "natural stream" was actually a "constructed conveyance" did not exist, urbanized, etc.).

Humboldt River, Nevada: This project is just getting started and will also use RF3 data, which is now being QA'd.

Region IX indicated that it would be helpful to have links to relevant data on the EMAP Home Page, since it is often difficult to locate and obtain pre-existing data, including older EMAP data and STORET data. Any link that can connect the user with a data contact would be helpful, along with any improvement in actually obtaining the data once it is requested.

Region IX also suggested that it would be helpful to have more imagery and land use available during the design phase of a study, in order to understand the landscape before field work is initiated (e.g., what landscapes you are working with, where water is found, agricultural use types). MRLC will be useful, although the scale is too coarse for working with river reaches and site information. For example, in the Central Valley project, they used Department of Agriculture aerial photos to find areas with water. USGS/Menlo Park is also developing a 1:48,000 digital orthophoto set that will be helpful.

Data Management
Data for the Southern California Bight was managed at Region IX, but the data has been transferred to the Southern California Coastal Water Research Project (SCCWRP) organization for distribution and archiving on the SCCWRP web site, which is linked to the EMAP Public Web Site (EMAP 1998).

Data entry for the Central Valley and Humboldt River projects is accomplished through automated data collection programs on laptops provided by ORD-Corvallis, and the data will be returned to Corvallis for archiving and distribution.

Data Quality Assurance/Quality Control
Data QA/QC is done by project managers, as follows:

- Southern California Bight: Larry Cooper, SCCWRP;
- Central Valley, California: Bob Hall, U.S. EPA Region IX, Mike McDowell, EPA Contractor, Corvallis; and
- Humboldt River, Nevada: will be done by project staff at University of Nevada, Reno.

Data Products

- Southern California Bight: The project produced a field data collection protocol modified from a protocol provided by ORD-Atlantic Ecology Division that forced researchers to collect data uniformly and reduced data entry errors. They implemented this protocol because each data collector was entering data differently on the cruises. To find out more about this modified protocol, contact SCCWRP 714-894-2222, or Bob Hall.
- Central Valley, California: No supplemental data were created, but they are working with California Fish and Game on a protocol for non-wadeable streams that will be available soon.
- Humboldt River, Nevada: No data collected yet. They will be testing new indicators and adapting EMAP protocols for arid environments.

Data Distribution
Data will be distributed by the site most appropriate for each data set.

- Southern California Bight: Data are now being distributed on the SCCWRP Web Site (SCCWRP 1998).
- Central Valley, California: Data will be given to ORD-Corvallis, who will decide on the most appropriate repository. Mike McDowell is the lead for making this data available. All of the data are publicly available, except for the names and addresses of landowners, which were removed from the database. Latitude and longitude of ditch sampling sites are still in the data.
- Humboldt River: These data will be given to ORD-Corvallis, who will decide on the most appropriate repository.

Documentation
Metadata is prepared by Region IX. SCCWRP metadata is already on their web site. Central Valley metadata will be prepared when the data analysis is completed.

Software, Hardware, Infrastructure
Region IX has Arc/Info, Access, and other packages, but often does not have software and hardware that can run ORD models (e.g., SAS routines). As a consequence, they have to make major modifications (e.g., re-programming) or cannot use the model. This situation is inefficient from a program perspective and Regions often do not have sufficient resources in R-EMAP budgets to make the modifications or obtain the appropriate software. Money for R-EMAP/EMAP software and hardware support has previously been allocated in OIRM budgets and should be released to the Regions.

Region IX has adequate Internet access and computing resources to manage and store R-EMAP data files and to access the EMAP Public Web Site.

Project Management
The R-EMAP coordinator is involved in all aspects of the projects, coordinating the researchers who include Publicly Owned Treatment Works (POTWs), academic researchers, state agencies, and others. The R-EMAP coordinator is also directly involved in ensuring the usability and delivery of data.

How Can EMAP–IM Help?
Region IX agreed that the Data Directory will be useful if it can help in locating and obtaining relevant data, including EMAP, STORET, landscape, and regional data sets. The Data Directory should have threads that lead users to related information of interest. The Region asked that the Data Directory be a cross-referenced directory to guide researchers to useful ORD indicators. For example, if a project found a certain organism during sampling, the directory could guide them to data that would indicate to what the organism is susceptible, of what it is an indicator, etc.

Problems that need to be resolved include minimizing the effort required to use baseline data and EMAP tools:

- the Region needs assistance with obtaining land use/landscape imagery data before/ during project design phase.
- the Region needs assistance with locating and obtaining existing ORD, EMAP, and other Federal data. The Data Directory will help, but they will need help obtaining the data.
- OIRM software and hardware funds should be released to the Regions as planned.

B.8 Ecological Indicator Development

B.8.1 Ecological Indicator Development Guidelines and Documentation

Summary of Conference Call
March 31, 1998

Participants

ORD-Gulf Ecology Division
- Bill Fisher

EMAP–IM (AED)
- Stephen Hale

Technology Planning and Management Corporation, ITAS Contractor
- Jeff Rosen
- Dillon Scott

Mission and Goals
The mission of this Working Group is to oversee the application of standards to development and evaluation of indicators (measurable or calculated parameters) according to standards that will ensure robust and diagnostic tools that can help define the status of critical environmental resources. The program to develop and evaluate indicators is an ORD-wide initiative, and the role of the Working Group for EMAP will be to coordinate research and oversee evaluation of indicators being developed by other Working Groups such as Landscape Ecology, MAIA–Surface Waters, and MAIA–Estuaries.

The Working Group's Draft *Research Strategy* (U.S. EPA, 1997c) outlines four primary goals:

- Identify indicator priorities and assessment endpoints, 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 (Program Offices) and users (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 developing risk assessment endpoints. However, the research will also strive to anticipate management needs for indicators of ecosystem integrity and sustainability. Several indicators will be developed and evaluated each year. The research will be conducted by a combination of ORD researchers and grant-funded academic researchers (through the EPA STAR grants program).

The principal activity of this Working Group in carrying out its mission will be to develop a framework for evaluating and documenting indicators. This process will consist of an evaluation of the indicator according to *Ecological Indicator Evaluation Guidelines* and a formal peer review by a committee established by the Working Group.

The *Ecological Indicator Evaluation Guidelines* now being developed by Laura Jackson for the Working Group will define 4 phases of indicator evaluation which include 15 guidelines:

- Phase 1. Conceptual Relevance, including two guidelines:
  - ❑ Relevance to an identified assessment
  - ❑ Relevance to ecological function
- Phase 2. Feasibility of Implementation, including five guidelines:
  - ❑ Data collection methods
  - ❑ Logistics
  - ❑ Information management
  - ❑ Quality assurance
  - ❑ Monetary costs
- Phase 3. Response Variability, including five guidelines:
  - ❑ Estimation of error
  - ❑ Temporal variability within field season
  - ❑ Temporal variability across years
  - ❑ Spatial variability
  - ❑ Discriminatory variability
- Phase 4. Interpretation and Utility, including three guidelines:
  - ❑ Power of detection
  - ❑ Assessment thresholds
  - ❑ Linkage to management actions

Each indicator must pass sequentially through the phases and cannot be evaluated under a later phase until it has passed through the previous phase.

For any indicator that meets the *Guidelines*, the Working Group will set up an indicator evaluation panel of scientific experts and risk assessors who can determine the strengths and weaknesses of the indicator. The peer review process will:

- 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.

Data Collection, QA/QC, Analysis, Aggregation
QA/QC of indicators will consist of the evaluation by the Guidelines and the peer review process.

Data will only be collected on an as-needed basis by researchers conducting pilot studies to test their indicators. Researchers will conduct their own data management, QA/QC, and analysis. The Working Group does not currently have a management plan or standards for handling the data. It is unlikely that data aggregates will be produced during indicator development and review.

Data Distribution
For information about how the indicator documentation will be distributed, see the Documentation subsection, below.

Monitoring data used to test indicators will remain with the researchers who developed it, and there are no plans for distributing the data sets. However, the data will be cited, documented, and given to peer reviewers.

Documentation
The principal product of this group will be the *Guidelines* and the text documentation of the indicator evaluation and peer review processes. EMAP–IM will track the documentation, which will consist of the following text-based products:

- The *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 (for the purpose for which it was designed); and
- Links and references to any data sources used to demonstrate and justify indicators under the *Guidelines*.

Data Sources
Previously collected data sources will be used to support the indicator presentation to the peer review panel. There will be a large variety of data sources, since indicators are being developed in a number of selected areas of importance to EMAP:

- Forests;
- Fresh Waters (lakes, streams, wetlands);
- Estuaries and coastal wetlands;
- Landscapes (across resource types); and
- Integration of whole ecosystem.

Data sources will be obtained by researchers and are expected to come from manuscripts, field notes, published literature, historic databases, and pilot monitoring programs. Data sources will be cited in the indicator documentation.

Data Volumes
The volume of this text-based data will probably be in the 10's to 100's of MB.

Users
Users of the documentation will be other EMAP programs that want to use the evaluated indicators and the Guidelines.

Exchange of data and information among indicator researchers is infrequent and takes place by direct contact between the individuals, so they do not see a need to index data sources or provide a web site for data exchange.

Software, Hardware, Infrastructure
The software used by the Working Group is not relevant to this analysis as long as they provide products to EMAP–IM in accessible formats for posting on the EMAP Public Web Site.

Project Management

The project management structure of this Working Group is designed encourage participation by both ORD clients and users in all stages of the indicator development and evaluation process, including reviewing all documents produced by the Working Group (e.g., *Research Strategy, Evaluation Guidelines*). This participation will be implemented in the following ways:

1. **Working Group Membership.** Because the Ecological Indicators research combines a broad spectrum of ORD's ecological research activities, the membership of the Working Group includes representatives from the National Exposure Research Laboratory (Cincinnati), National Health and Environmental Effects Research Laboratory (Research Triangle Park), National Center for Environmental Assessment (Washington, DC), and the National Risk Management Research Laboratory (Ada, OK). In addition, the Working Group includes representation from Office of Research and Science Integration to assist in responsiveness to client needs, and National Center for Environmental Research and Quality Assurance to assist in integrating the research of academic scientists through grant support.

2. **Peer Review Panel Membership**. The peer review panels will be made up of a number of outside scientists who are potential users of the indicators. In order to develop the panels, the Working Group is developing a formal Client Liaison Group that will bring the needs of clients and environmental managers into the peer review process in order to ensure that indicators are applicable to ongoing monitoring programs. The Client Liaison Group will include state and federal risk assessors, who will be able to influence indicator evaluation by their participation on the indicator evaluation panels. The Working Group is trying to get the broadest possible base of participation, and have recently been conducting outreach to get participation by all EPA Regions and program offices, states, private industry (e.g., DuPont, the Chemical Manufacturers Association), and others. Participants can choose where they want to provide input, and, therefore, the makeup of peer review panels will vary based on the content of the indicator being reviewed. The group contains members from all NERL labs; some participants are very active, others only participate on issues of special interest to them.

   The efforts of this Working Group are based on consensus—including submission of indicators for review, and participation by reviewers. The original charter of the Working Group was to assemble people interested in developing indicators; the evaluation process grew out of that and has now included developing the *Guidelines* and *Strategy*. The Working Group is not sure if the *Guidelines* will be used by other organizations, but they are beginning to apply it within ORD. They have begun evaluating sample indicators, and in the future will be evaluating MAIA indicators (Landscape Ecology, Surface Waters, Estuaries). Three sample indicators have already been completed and are being evaluated using the *Guidelines*—Laura Jackson is directing development of the *Guidelines* and the sample

indicators, which are in draft stage. The process of developing and evaluating indicators will proceed over a long period of time as new indicators are developed by researchers and submitted for review. Indicator reviews will be initiated as needed. They are now getting assistance from NERL senior managers Rick Linthurst and Kate Smith, who will write an overview and fill in some gaps in the *Research Strategy*. The Working Group now also has a cooperative agreement with the National Research Council covering criteria and guidelines for good indicators.

EMAP–IM Assistance

EMAP–IM can assist by distributing this Working Group's documentation of the review process on the EMAP Public Web Site, including the *Guidelines*, indicator descriptions, peer review summaries, and links to or citations of supporting data.

EMAP–IM can also help by creating links on the EMAP Public Web Site to web sites where non-EMAP indicators and their development are referenced.

In the future, EMAP–IM could assist in developing a tracking system (e.g., classification scheme for indicators, events the Working Group needs to track, information about indicator development by other agencies) similar to the Data Directory that would indicate the status, evaluation, and peer review of each indicator and index indicators by type and resource so users can search by area of interest. However, the Working Group feels that there are not yet enough indicators being evaluated to justify a tracking and indexing system.

The most important activity of this Working Group and EMAP–IM is to document and distribute the results of the indicator evaluation process so that users can understand the quality and applicability of ecological indicators.

B.8.2 Aquatic Mortality Monitoring Database

EMAP will assist states with developing a database that incorporates tracking information about mortalities of marine and estuarine aquatic organisms. Mortality events are important not only because of the loss of affected organisms, but also because they may signal the presence of public health dangers or degrading 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 condition at different spatial scales.

This effort will be 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 goal of documenting and understanding the causes of epizootic and mortality events in the Gulf of Mexico.

Goals of the program are (GMNET 1998):

- Improve interstate communication among mortality response teams: this will 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.

EMAP–IM (AED) will 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 cause(s) of mortality events.

This effort will involve:

- establishing communication among the states' efforts;
- merging and integrating mortality information to meet the common goal 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 Geographic Information System (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 in order 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 result will be:

- high-quality regional data;
- improved reporting of mortality 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.

B.9 Committee on the Environment and Natural Resources

Stephen Hale spoke with Tom Mace (Chair, CENR Data Management Working Group) on December 17, 1997, and Jeff Rosen spoke with him on May 11, 1998. The needs related to EMAP's coordination with CENR are considered throughout the Information Management Plan, especially in Sections 1.3 (EMAP Participation in the Committee on the Environment and Natural Resources) and 5.5.2.2 (Coordination with the Federal Interagency Committee on the Environment and Natural Resources). CENR will adopt data standards and protocols for documenting data and facilitating exchange, including the Global Change Research Program's Global Change Data and Information System (GCDIS).

B.9.1 CENR Information

This section reproduces a text file provided by Tom Mace and contains detailed information about CENR activities.

Data Management for Global Change Research
Policy Statements for the National Assessment Program
Subcommittee on Global Change Research— May 8, 1998

The overall purpose of these policy statements is to facilitate full and open access to known quality data and information resulting from global change research assessments for their use with confidence both now and in the future. These policies reflect the goals and policies of the U.S. Global Change Research Program and incorporate federal laws, directives, and regulations regarding the maintenance and dissemination of data and information in the Federal Government. These policies are recommended for all participants in the National Assessment Program, including not only federal agencies but state, local, tribal, foreign, educational, non-government organizations, and private partners.

- The U.S. Global Change Research Program's National Assessment Program requires an early and continuing commitment to the establishment, maintenance, description, accessibility, and distribution of high-quality, data and information.
- Full and open sharing of the full suite of data and published information produced by the Assessment Program is a fundamental objective. Data and information should be available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution.
- There should be archival and documentation of all data sets used in and resulting from the National Assessment Program. At a minimum, the documentation for each data set should include a quality assessment and a citation containing the six required fields for the Directory Interchange Format. Both should be made accessible over the Internet with ANSI Z39.50 search and retrieval standard compatibility.
- Standards used for the descriptions of individual data sets should be compatible with those of the Global Change Master Directory. Spatial data set descriptions should also be compatible with the Content Standards for Digital Geospatial Metadata of the Federal Geographic Data Committee. All data set documentation and descriptions should be made available for inclusion in the Global Change Master Directory.
- Contributors to the Assessment Program should actively participate its Web page to share information and coordinate the Program's disparate activities.

Data Management for Global Change Research Policy Statements
Executive Office of the President, OSTP - July 2, 1991

The overall purpose of these policy statements is to facilitate full and open access to quality data for global change research. They were prepared in consonance with the goal of the U.S. Global Change

Research Program and represent the U.S. Government's position on the access to global change research data.

- The U.S. Global Change Research Program 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 full 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 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 arrangements for exchanging data among researchers.
- For those programs in which selected principal investigators have initial periods of exclusive data use, data should be made openly available as soon as they become widely useful. In each case, the funding agency should explicitly define the duration of any exclusive use period.

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. *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.

b. *OMB Bulletin 95–01, Establishment of Government Information Locator Service (GILS)*, December 7, 1994, is designed to help the public and agencies locate and access information electronically throughout the U.S. government.

c. *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.

d. *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.

e. *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.

f. *Executive Order 12862, Setting Customer Service Standards*, September 11, 1993, mandates easy accessibility of federal government information and services.

g. *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.

h. *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.

i. *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.

j. *44 United States Code Chapters 17 and 19* define the legal requirements for providing information to the public through the Federal Depository Library Program.

k. *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.

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.

Table B-3. Requirements Analysis Questions from Landscape Ecology Interview

A set of requirements analysis interview questions as prepared for each Working Group. The questions prepared for Landscape Ecology Working Group are presented here as an example.

Landscape Ecology Presentation
The purpose of the meeting is:

- To understand and document Landscape Ecology's needs for information management;
- To gather the information necessary to revise and update the Draft Information Management (IM) Plan;
- To evaluate the appropriateness of the Draft IM Plan to meet Landscape Ecology and EMAP information goals;
- To address Red Team Review unresolved issues of data requirements and project management; and
- Not to dictate information management policy or approach within Landscape Ecology.

The following data requirements questions were posed to the Landscape Ecology Working Group:

Mission & Goals
What are the mission and goals of Landscape Ecology?
Does the Landscape Ecology group have additional support documentation?

Research Plan
What are the major data generating projects planned?
Is the same research approach being used for all geographic regions?
In MAIA region, what geographic area has been completed and what will ultimately be completed?
What is an "indicator coverage"?

Purpose
Who are the primary recipients of analyses and data aggregates (e.g., decision makers, policy staff)?
What are the benefits of the data and analyses done by Landscape Ecology?

Data Sources
What data sources will be accessed and used?
What formats are these data sources in?
What are the expected uses, development, or aggregation of each data source?

Data Sources
What is the perceived quality and scale of these data sources?
How often are these sources accessed?
How often are they updated?
Who are the stewards of the data sources?
What are the longevity and quality of this stewardship?
What is the adequacy and accessibility of documentation of data sources?
How easily accessible are the data sources—what are the barriers?
What EMAP data are useful for Landscape Ecology?
What data types will Landscape Ecology need to exchange with other EMAP groups?
What groups collaborate with which Landscape Ecology projects?
How much effort is expended tracking down needed data sources?
How could the EMAP Data Directory or other tools be helpful in finding needed data sources?
What are the future data source needs of the Landscape Ecology group?

Data Collection
What data types will be collected and stored?
What will be the volume of these data?
What format will the data be collected and stored in?
What is collection schedule of each data type (number of sampling events, intervals, etc.)?
What data sets will be produced in the future?

Data Aggregation
What are the expected data aggregation products from the research?
What tools, analyses, and processes are used to produce these aggregates?
Can these data aggregates be shared?
Whom can/should they be shared with?

Data Gaps
What major data gaps need to be filled?
How can these data gaps best be filled?

Data Products
What types of data outputs are generated for each project (tables, GIS coverages, etc.)?
What formats/content of data products will be made available from each project?
What projection will coverages be in? vector/raster? what scale?

Users
Who are the primary data users, and what do they do with data?
Who are the secondary data users, and what do they do with data
What other users should be considered?

Data Management
How are data managed?
What is the data flow?

- data entry
- verification
- validation
- QA/QC
- security
- access
- storage

What databases and data structures are used for data management?
What data standards are used in Landscape Ecology?
What data standards would Landscape Ecology like to see EMAP adopt or recommend to other EMAP groups?
Can EMAP–IM assist with data management, or data and metadata standards?
What is the physical location of data sets?
Where will analyses be performed?

Data Flows
What are the data flows?

- data collection (internal or external source) to data entry
- data entry to QA/QC
- QA/QC to analysis
- analysis to aggregation
- aggregation to reporting
- reporting to delivery
- delivery to distribution
- archiving
- updating
- updating to redistribution (users)

How are data exchanged between collaborators?
How are links made to other relevant data or provided for EMAP-L data?

Metadata
What are the minimum requirements/standards for documentation of collected data and data aggregates for use within Landscape Ecology? Outside users (non-EMAP scientists, general public, etc.)?
How are these requirements compatible with EMAP Data Catalog or FGDC?
Who is responsible for creating and making metadata accessible?

Data Access/Distribution
Which internal and external users will want data sets? How will they gain access to them?
How will preliminary data (before full QA and analysis) be made accessible to EMAP researchers?
What other EMAP and non-EMAP groups need data from Landscape Ecology?
How will data sets be distributed to them?
How can Landscape Ecology ensure long-term access by EMAP to the data sets?
What kinds of data sharing vehicles are anticipated (e.g., EMAP Public Access Home Page, CD-ROM)?
Is there a need for EMAP–IM support to facilitate data exchanges?
What security and data preparation issues prevent sharing and distribution of data sets, such as issues of security, timing, or money?
What incentives can facilitate sharing and documentation of data within EMAP and between EMAP and its partners?

Functionality
What staff are responsible for data management?
What is their allocation (full-time/part-time/etc.)?
What level of support is required?
What level of support is desired?
What software/hardware configuration is are used for accessing, storing, exchanging, and analyzing data?
What software/hardware configuration is required?
What other features are desired?
What network configuration and operating systems are available for meeting needs of data access, analysis, and exchange?
What configuration is required?
What other features are desired?
What Internet access and bandwidth are available? Are they adequate to support information needs?
What media are required?

What media would be desirable?
What media exist for exchange and access of internal and external data sources (tape, Internet, client/server)?
What media are required?
What media would be desirable?
What interfaces are available to display and download metadata, tabular and spatial data to/from other data systems?
What interfaces would be required?
What features of interfaces would be desirable?
What plans and funding exist for upgrading existing functionality to adapt to technological change and/or to implement the required or desired functionalities discussed above?

Project Management/Administration
What is the project management structure of the Landscape Ecology group?
What is the EMAP or Landscape Ecology project management relationship with cooperating groups?
Who is ultimately responsible for Landscape Ecology data quality and maintenance?
Who is the Landscape Ecology liaison with EMAP–IM?
Who decides about data flows, release, security?
What are the key functions of staff that generate, analyze, manage, use data?

Data Directory
Who in the research group or ORD will create and update Data Directory entries?
How can EMAP–IM assist with transfer of entries to the EMAP Data Directory?
What does Landscape Ecology need from EMAP–IM?

- Access to new images and data from download sites, and processed images with appropriate documentation
- Access to ground-truthing data
- Directory of and access to available 1990–1995 data sets
- Access to other data specified in Research Plan

Role of EMAP–IM
How would a directory or clearinghouse help Landscape Ecology's efforts?
What would make it easier for Landscape Ecology to participate actively in the EMAP–IM system?
Should EMAP–IM store and manage data sets of interest to EMAP that have no long-term steward and incorporate them into database, Directory, and Catalog?

What EMAP–IM needs from Landscape Ecology
Access to EMAP data and metadata

- Directory entry information for data sets
- Metadata for each data set
- Bibliography
- Electronic publications
- Designated liaison

Meeting Summary
Summarize information heard during meeting for preliminary approval by participants

- What is the IM Plan for Landscape Ecology?
- How well does EMAP–IM plan support Landscape Ecology's objectives?
- Does the current EMAP–IM project management approach give Landscape Ecology access to the data and resources it needs?
- Are Landscape Ecology and EMAP–IM approaches flexible enough to meet user needs and evolve with changing technologies?
- How will they fare given pertinent management issues and existing IM frameworks?

Next Steps

- Prepare summary of this EMAP research group meeting and send it to participants for review
- Interview other EMAP research groups about their data and system needs
- Revise IM Plan in accordance with EMAP research groups feedback and system needs
- Distribute revised IM Plan for review
- Prepare final IM Plan in response to review
- Begin IM planning for 1998–2001

Appendix C
Inventory of EMAP Data

C.1. Purpose
C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data
C.3 Types, Volumes, and Status of Current EMAP (1996-) Working Group Data
C.4 Types, Volumes, and Status of Other Data

C.1 Purpose

This document fulfills part 3 of the EEI-1 documentation requirements by providing a information about EMAP data types (see Appendix A, EEI Requirements Report). This section presents a high-level inventory of data types, estimated volumes, status, and repositories. The inventory of Resource Group 1990–1995 data is very detailed because the projects are completed and all of the data resides in ORD or with close collaborators. Data for the 1996–2001 Working Groups are less well known because most of the projects are in their early stages and much of the data is being collected in partner agencies in many different formats, not all currently specified or known to EMAP–IM. Resource Group was obtained from the EMAP Public Web Site (EMAP 1998) and from an interview with the Forests Resource Group data manager. Working Group information was obtained during Requirements Analysis interviews and information provided by the research groups. The level of information provided in this appendix is the most detailed available at this writing (June 1998) and is adequate to allow EMAP to design the IM system to anticipate major data exchange and storage needs because most of the data will not be stored in the EMAP system. More information about Working Group data products will be added in future versions of the Plan as it becomes known.

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data

Resource Group	Study Type and Duration	Data Set Type and Format	Approximate Data Volumes	Repository and Status
Estuaries **Virginian Province**	Assessment of status and trends in condition of estuaries extending inland to the head of tide (including coastal embayments, bays, inland waterways, tidal rivers, coastal wetland areas and salt-water marshes) from Virginia to Massachusetts. 1990–1993	Station Location (ASCII)	1990-111 records-7127 bytes 1991-101 records-6518 bytes 1992-102 records-6644 bytes 1993-110 records-7062 bytes	EMAP Public Web Site Available
		Station Information (ASCII)	1990-111 records-9298 bytes 1991-101 records-8508 bytes 1992-103 records-8666 bytes 1993-110 records-9219 bytes	
		Events (ASCII)	1990-111 records-9607 bytes 1991-101 records-8827 bytes 1992-102 records-8983 bytes 1993-110 records-9529 bytes	
		Vertical profile water quality (ASCII) Surface		
		Bottom	1990-111 records-8862 bytes 1991-101 records-8143 bytes 1992-102 records-8586 bytes 1993-110 records-8790 bytes	
		QA Table	1990-2 records-488 bytes 1991-9 records-812 bytes 1992-16 records-1534 bytes	
		Sediment toxicity tests (ASCII)	1990-111 records-9024 bytes 1991-101 records-8264 bytes 1992-102 records-8416 bytes 1993-110 records-8948 bytes	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Sediment chemistry (ASCII)	1990-4675 records-407KB 1991-7940 records-706KB 1992-7669 records-684KB 1993-9064 records-834KB	
		Fish species (ASCII)	1990-440 records-34KB 1991-477 records-38KB 1992-468 records-38KB 1993-537 records-43KB	
		Fish community (ASCII)	1990-111 records-4823 bytes 1991-101 records-4433 bytes 1992-103 records-4511 bytes 1993-110 records-4784 bytes	
		Benthic community (ASCII)	1990-111 records-11205 bytes 1991-101 records-10275 bytes 1992-103 records-10461 bytes 1993-110 records-1112 bytes	
		Benthic species (ASCII)	1990-3515 records-27KB 1991-3342 records-268KB 1992-3090 records-239KB 1993-3288 records-263KB	
		Benthic replicate abundance (ASCII)		
		Benthic replicate biomass (ASCII)		

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Sediment grain size composition (ASCII)	1990-111 records-9212 bytes 1991-101 records-8442 bytes 1992-103 records-8596 bytes 1993-110 records-9135 bytes	
		Tissue chemistry (1991 only) (ASCII)	1991-1887 records-151KB	
		Virginian province sampling station locations (ARC/INFO export)	40KB	
		Watersheds for Virginian Province small estuaries (ARC/INFO export)		

Resource Group	Study Type and Duration	Data Set Type and Format	Approximate Data Volumes	Repository and Status
Estuaries **Carolinian Province**	EPA and NOAA joint assessment of status and trends in condition of estuaries extending inland to the head of tide (including coastal embayments, bays, inland waterways, tidal rivers, coastal wetland areas and salt-water marshes) from North Carolina to Florida. 1994--1997	Stations (ASCII)	342 records-27KB	Summary Data: ORD Atlantic Ecology Division and EMAP Public Web Site
		Vertical profile water quality Surface	313 records-15KB	
		Bottom	313 records-16KB	
		Sediment toxicity tests (ASCII)	1272 records-76KB	
		Sediment chemistry (ASCII)	34,144 records-1.4MB	
		Demersal species abundance	2506 records-81KB	
		Demersal species lengths	1682 records-64KB	
		Demersal community (ASCII)	200 records-8KB	
		Benthic community (ASCII)	318 records-11KB	
		Benthic species abundance (ASCII)	5087 records-188KB	
		Benthic replicate abundance (ASCII)	7710 records-235KB	
		Sediment grain size composition	318 records-12KB	
		Tissue chemistry (ASCII)	2635 records-139KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued

Resource Group	Study Type and Duration	Data Set Type and Format	Approximate Data Volumes	Repository and Status
Estuaries **Louisianian Province**	Assessment of status and trends in condition of estuaries extending inland to the head of tide (including coastal embayments, bays, inland waterways, tidal rivers, coastal wetland areas and salt-water marshes) from Florida to Texas. 1991–1994 (1994 data not yet available)	Station Location (ASCII)	1991-100 records-7KB 1992-110 records-6KB 1993-100 records-1KB	Summary Data: ORD Atlantic Ecology Division and EMAP Public Web Site
		Station Information (ASCII)	1991-100 records-9KB 1992-110 records-10KB 1993-100 records-2KB	
		Vertical profile water quality (ASCII) Surface Water	1991-100 records-8KB 1992-110 records-8KB 1993-100 records-7KB	
		Bottom water	1991-100 records-9KB 1992-110 records-9KB 1993-100 records-8KB	
		Sediment chemistry (ASCII)	1991-10,800 records-824KB 1992-11,880 records-906KB 1993-10,200 records-767KB	
		Sediment analyte code (ASCII)	1991-11KB	
		Sediment toxicity (ASCII)	1991-200 records-16KB 1992-220 records-15KB 1993-200 records-15KB	
		Sediment grain size composition (ASCII)	1991-100 records-6KB 1992-110 records-5KB 1993-100 records-5KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Benthic species abundance (ASCII)	1991-1288 records-182KB 1992-3145 records-250KB 1993-32880 records-257KB	
		Benthic community (ASCII)	1991-100 records-6KB 1992-110 records-7KB 1993-100 records-5KB	
		Demersal abundance by trawl (ASCII)	1990-499 records-41KB 1991-633 records-51KB 1992-692 records-54 KB	
		Demersal community (ASCII)	1991-100 records-6KB 1992-110 records-7KB 1993-100 records-6KB	
		Tissue chemistry (ASCII)	1991-100 records-7KB 1992-110 records-8KB 1993-100 records-6KB	
		Tissue chemistry codes	1991-6812 records-526KB 1992-10,890 records-852KB 1993-7007 records-541KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Surface Waters	Study of all U.S. inland lakes (excluding the Great Lakes) and streams. Lakes and streams in the sampling grid were classified by size, length, geographic location, elevation, and ecoregion. Samples were obtained for each lake and stream class to estimate ecological condition. Mid-Atlantic Streams 1993-1994 Northeast Lakes 1991-1994	Mid-Atlantic Streams Benthic counts (ASCII) Benthic metrics (ASCII) Water chemistry (ASCII) Fish counts (ASCII) Fish metrics (ASCII) Fish names (ASCII) Fish tissue contaminants (metals) summarized by stream (ASCII) Fish tissue contaminants (organics) summarized by stream (ASCII) Watershed characteristics summarized by stream (ASCII) Habitat summarized by stream (ASCII) Site information design (ASCII) *Region III R-EMAP mid-Atlantic highlands data set holds related data for mid-Atlantic streams*	 27,868 records-3.4MB 680 records-218KB 494 records-182K 2705 records-204KB 303 records-62KB 253 records-16KB 136 records-27KB 136 records-58KB 448 records-136KB 189 records-191-KB 588 records-111KB	Summary Data:

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Northeast Lakes		
		Breeding bird count (ASCII)	8960 records-735KB	
		Breeding bird metrics (ASCII)	307 records-163KB	
		Breeding bird names (ASCII)	217 records-18KB	
		Water chemistry (ASCII)	500 records-174KB	
		Sediment diatom counts by lake (ASCII)	34208 records-4.4MB	
		Sediment diatom metrics summarized by lake (ASCII)	730 records-87KB	
		Fish count summarized by lake (ASCII)	12302 records-1.6MB	
		Fish metrics summarized by lake (ASCII)	264 records-60KB	
		Fish names (ASCII)	90 records-6KB	
		Fish tissue contaminants (metals) summarized by lake (ASCII)	256 records-62KB	
		Fish tissue contaminants (organics) summarized by lake (ASCII)	178 records-73KB	
		Watershed characteristics summarized by lake (ASCII)	367 records-72KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Zooplankton counts summarized by lake (ASCII)	18259 records-1.5MB	
		Zooplankton metrics summarized by lake (ASCII)	593 records-128KB	
		Zooplankton names (ASCII)	392 records-13KB	
		Site information design (ASCII)	500 records-95KB	
		Northeast sampled lakes and watershed boundaries (ARC/INFO export)	9MB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Landscape Characterization	Used remote sensing and geographic information systems to estimate the areal extent of various kinds of resources within a given region. Assisted other EMAP Resource Groups with defining sampling frames by specifically locating and mapping the area of all resources of a particular kind within a region of interest.	CERCLA sites	356,261 bytes	Summary Data: EMAP Public Web Site
		National Priority List (NPL) hazardous waste sites	539,725 bytes	
		Digital Elevation Model (DEM) for Region III	Not available	
		U.S. Hydrologic Units (HUC) for Region III	594,771 bytes	
		Watershed boundaries for Region III		
		U.S. forest cover types		
		Forest cover types for Region III	456,293 bytes	
		RCRA treatment, storage, & disposal sites in Region III	63,319 (Compressed Arc/Info Export format)	
		Toxic Chemical Release Inventory for Region III		
		Wash., D.C.	9334 bytes	
		DE	204,577 bytes	
		MD	581,139 bytes	
		PA	3,274,165 bytes (3.2MB)	
		VA	1,077,167 bytes (1 MB)	
		WV	472,911 bytes	
		Omernik's Ecoregions for Region III	378,709 bytes	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Omernik's Subecoregions for Region III	597,291 bytes	
		Railroads for Region III (1:2,000,000)	604,183 bytes	
		Roads for Region III (1:2,000,000)	874,051 bytes	
		Hydrologic features for Region III (1:2,000,000)	1,064,667 bytes	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Forests	EMAP's forest sampling monitored indicators of forest resource condition on all forested sample sites defined on the Forest Health Monitoring (FHM) monitored indicators of forest resource condition on all forested sample sites defined on the EMAP systematic grid. The study currently includes samples from 27 states*. The research is conducted by U.S. Forest Service's National Forest Health Monitoring Program (U.S. Department of Agriculture). EMAP is a cooperator in this program and will use the FHM data as the core source of forest information. Data collection continuing from 1998 on. *States include: Alabama, California, Colorado, Connecticut, Delaware, Georgia, Idaho, Indiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming	1990–1997 Summary Data—*Data are for all states sampled.* Plot-level summaries (ASCII, SAS) Tree-level summaries (ASCII, SAS) Data dictionary for variables Code list for variables Condition codes Bio-indicator plants Regional and national data aggregates will be produced for MAIA researchers by FHM with a mensuration engine (SAS data analysis tool to reduce plot and tree summaries) Metadata: Online documentation in Oracle database (not yet adequate for non-forestry experts) and documentation of methods in "Forest Health Monitoring Field Methods Guide"	Total approximately 70 MB	Summary Data: Maintained by Forest Health Monitoring (FHM) information management team, U.S. Forest Service, Las Vegas, NV. Some data aggregates will be delivered to MAIA researchers. Primary access will be through the future FHM web site (http://www.fs.fed.gov). Hyperlinks to this site will be established through the EMAP Public Web Site.

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Agroecosystems	Nebraska Pilot—Results of questionnaire of farm operators by USDA-National Agricultural Statistics Service Nov–Dec 1993	cropping sample frame criteria farm field metrics crop & field management histories field operations irrigation practices pest management activities	SAS	Summary Data: held by North Carolina State University for USDA Not yet available

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Great Lakes	The Great Lakes resource comprises the five Great Lakes (Superior, Michigan, Huron, Erie, Ontario), including river mouths up to the maximum extent of lake influence; wetlands contiguous to the lakes; and the connecting channels, Lake St. Clair and the upper portion of the St. Lawrence Seaway.	Lake Michigan (1994)		
		Sediment data		
		Toxicity test data	78 records-5KB	
		Sediment quality data		
		Benthos data	636 records-51KB	
		Particle Size data	35 records-2KB	
		Station data		
		Location data	39 records-1KB	
		Water Quality data		
		Chlorophyll data	39 records-2KB	
		Total Suspended Solids data	39 records-2KB	
		CTD data	1218 records-55KB	
		Lake Ontario (1994)		
		Sediment data		
		Toxicity test data	106 records-6KB	
		Sediment quality data		
		Benthos data	511 records-40KB	
		Particle Size data	52 records-3KB	
		Station data		
		Location data	58 records-2KB	
		Water Quality data		
		Chlorophyll data	160 records-12KB	
		CTD data	6185 records-429KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Lake Superior (1994)		
		Sediment data		
		Toxicity test data	48 records-3KB	
		Sediment quality data		
		Benthos data	235 records-17KB	
		Particle Size data	23 records-1KB	
		Station data		
		Location data	27 records-1KB	
		Water Quality data		
		Chlorophyll data	78 records-6KB	
		Total Suspended Solids data	27 records-1KB	
		CTD data	1445 records-118KB	

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Rangelands	The target population for rangeland sampling included the arid, semi-arid, and sub-humid regions of the conterminous United States. Rangelands include desert scrub, prairies, grasslands, chaparral, open woodland, alpine tundra, arctic desert, and riparian communities but excluded intensively managed agriculture such as irrigated farmlands. The research group used a hierarchical biogeographic classification system that allowed interpretations at multiple levels of biological organization. EMAP's Rangelands Group planned to collect data using both sample-based and remotely sensed information (e.g., satellite imagery) as part of the indicator research and monitoring strategy.			Data repository: U.S. EPA/ NERL-LV

C.2 Types, Volumes, and Status of Early EMAP (1990–1995) Resource Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Delaware & Maryland Coastal Bays	Collaborative effort among state & federal agencies to assess ecological condition of the system and fill a data void identified in previous characterization studies. Two hundred sample sites in probability-based stratified sampling design (includes Rehoboth Bay, Indian River Bay, Assawoman Bay, Chincoteague Bay, upper Indian River, St. Martin River, Trappe Creek, dead-end canals).	Stations (ASCII)	264 records-13KB	
		Station Information (ASCII)	204 records-8KB	
		Surface water quality (ASCII)	264 records-31KB	
		Bottom water quality (ASCII)	264 records-19KB	
		Benthic species abundance (ASCII)	5546 records-531KB	
		Benthic assessment (ASCII)	200 records-16KB	
		Fish species abundance (ASCII)	462 records-24KB	
		Sediment chemistry (ASCII)	3780 records-277KB	
	Summer 1993	Sediment chlorophyll (ASCII)	248 records-9KB	

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data

Resource Group	Study Type and Duration	Data Set Type and Format	Approximate Data Volumes	Repository and Status
MAIA ORD Regional Assessment–Surface Waters	Collecting 1000's of parameters from 150–200 stream stations every summer for 1993–1998 and continuing into the future.	Benthic counts (ASCII) Benthic metrics (ASCII) Water chemistry (ASCII) Fish counts (ASCII) Fish metrics (ASCII) Fish names (ASCII) Fish tissue contaminants (metals) summarized by stream (ASCII) Fish tissue contaminants (organics) summarized by stream (ASCII) Watershed characteristics summarized by stream (ASCII) Habitat summarized by stream (ASCII) Site information design (ASCII)		Western Ecology Division and EMAP Public Web Site Data still being analyzed and documented

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
MAIA ORD Regional Assessment–Estuaries	Collecting standard set of EMAP–Estuaries monitoring parameters for two years 1997B8 [including water quality parameters, benthic infauna, sediment toxicity (1997 only), and fish trawls (1998 only)]	Station Location Data (ASCII) Station Information Data (ASCII) Events (ASCII) Vertical profile water quality (ASCII) Surface Data Bottom Data QA Table Sediment toxicity tests (ASCII) Sediment chemistry (ASCII) Fish species (ASCII) Fish community (ASCII) Benthic community (ASCII) Benthic species (ASCII) Benthic replicate abundance (ASCII) Benthic replicate biomass (ASCII) Sediment grain size composition (ASCII) Tissue chemistry (1991 only) (ASCII)	Data still being analyzed and documented.	Atlantic Ecology Division and EMAP Public Web Site

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
MAIA ORD Regional Assessment–Landscape Ecology	Landscape Atlas for MAIA region and landscape indicator GIS coverages	Approximately .5 GB indicator data in Arc/Info export and ASCII files	Indicator data will be submitted in 1998	EMAP Public Web Site
Intensive Sites–Terrestrial UV-B	Air monitoring parameters at 12 National Park sites, and historical data for urban monitoring sites)	Main Parameters: UV-A, UV-B, Diffey-weighted UV (DUV), Direct Sun Ozone (DSO3), Zenith Sky Ozone (ZSO3), Direct Sun Nitrogen Dioxide (DSNO2), Zenith Sky Nitrogen Dioxide (ZSNO2), Irradiance	Data collection 1996 - ongoing Existing data approximately 300MB (National Park data only). Projecting approximately 1-2 GB per year for three years, depending on final number of sites chosen expected in UV-B system at NERL (estimated from May 1998 draft UVB Data Project, Overview and Options).	UV-B standalone database and EPA Environmental Information Management System (EIMS)
		National Park Sites Big Bend, Texas (1997)	Estimated volumes 53MB	In progress
		Great Smoky Mts, Tennessee (1996-1997)	71MB	
		Everglades, Florida (1997)	54MB	
		Shenandoah, Virginia (1996-1997)	62MB	
		Canyonlands, Utah (1997)	20MB	
		Glacier, Montana (1997)	16MB	
		Denali, Alaska (1997)	11MB	

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
		Olympic, Washington (1997)	5.6MB	
		Acadia, Maine (began operating 3/98)	N/A	
		Virgin Islands (not yet operating)	N/A	
		Rocky Mountain, Colorado (not yet operating)	N/A	
		Sequoia, California (future site)	N/A	
		Theodore Roosevelt, North Dakota (future site)	N/A	
		Hawaii Volcanoes, Hawaii (future site)	N/A	
Intensive Sites–Terrestrial Routine Air Monitoring	Data from existing routine monitoring stations in National Parks	Database maintained by National Park Service.	National Park Service data collection ongoing. No detailed information obtained on types and volumes.	National Park Service web sites
Intensive Sites–Terrestrial Research Grants	Data sets in a variety of disciplines from individually funded researchers (e.g., amphibian surveys, nutrient deposition and eutrophication)	Expecting moderately sized databases from 10's to 100's of MB (no data collected yet)	No grants issued yet	EMAP Public Web Site
Intensive Sites–Coastal	Monitoring at 41 sampling sites of a variety of disciplines (e.g., anthropogenic stressors; temporal & spatial variability of measurements; development of indicators; development of remote sensing capability)	Undetermined at this time	No projects funded yet	Undetermined at this time

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
Indicator Development	Principal product is text-based documentation, including: • Evaluation Guidelines • Summaries of the evaluation and peer review processes, and • Links to data sources used. Few monitoring data sets or data aggregates will be produced, and these will be maintained by researchers	10's to 100's MB of text-based information	Process just beginning (Guidelines in draft form, sample indicators being evaluated)	EMAP Public Web Site
Landscape Ecology	Landscape indicator GIS coverages, processed remote sensing images, landscape assessments.	Projected 250GB (in 1998)	Data analysis is in progress	EMAP Public Web Site
R-EMAP (see project-specific inventories, below)	Two to three monitoring projects in each of 10 EPA Regions.	10's to 100's MB for each project. Potential total of several GB overall R-EMAP, depending on amount of remote sensing data)	Ongoing	Variety of repositories including STORET, EMAP Public Web Site. See details in Appendix B, Data Management Needs and Practices of EMAP Working Groups.

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
R-EMAP Region I	Monitoring data sets for mercury in several areas: • fish tissue contamination, • mercury deposition and atmospheric concentrations, and • mercury in lake-bed sediments.	Currently on EMAP server: dBase/ ASCII data: 500 KB Metadata: 165 KB	Submitted in dBase, converted to ASCII	EMAP Public Web Site (temporary), STORET, Mercury Deposition Network, and others (Appendix B, Data Management Needs and Practices of EMAP Working Groups.)
R-EMAP Region II	Monitoring data for New York/New Jersey Harbors	All data sets in ASCII Station Location data Station Visit data Water Quality Surface data Water Quality Bottom data	 168 records-8KB 168 records-7KB 168 records-12.9KB 168 records-12.9KB	
		Benthic Species Replicate Abundance data Benthic Replicate Biomass data Sediment Chemistry data Sediment Analyte Code Resolution Amphipod Sediment Toxicity Test data Microtox Toxicity Test data Sediment Silt/Clay data	8330 records-891KB 5232 records-392KB 18480 records-1MB 110 records-6KB 168 records-9KB 168 records-11.9KB 168 records-5KB	

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
R-EMAP Region VI	Monitoring data for: • standard EMAP Estuaries monitoring parameters collected for Texas estuaries, and		Subset of whole database prepared for EMAP World Wide Web Site : SAS data: 2.04 MB Metadata: not prepared yet	
	• probabilistic approach to stream habitat degradation in east Texas streams	Sediment Chemistry	11996 records-1MB	EMAP Public Web Site
		Sediment Toxicity	166 records-14K	
		Sediment Grainsize	105 records-6K	
		Station Location	109 records-8K	
		Station Information	109 records-6K	
		Vertical Water Quality Surface Bottom	 109 records-7K 109 records-6K	
		Benthic Species Abundance	2481 records-189K	
		Benthic Community	115 records-6K	
		Fish Species Abundance Fish Tissue Chemistry	647 records-52K 5524 records-691K	
		Fish Species Codes	109 records-8K	

C.3 Types, Volumes, and Status of Current EMAP (1996–) Working Group Data, Continued				
Resource Group	**Study Type and Duration**	**Data Set Type and Format**	**Approximate Data Volumes**	**Repository and Status**
R-EMAP Region X - Southern Calif. Coastal Water Research Project	1977,1985, 1990,1994	1994 Pilot Project Data Benthic Data (ASCII)	169,147 bytes	SCCWRP Web Site (SCCWRP 1998)
		Sediment Chemistry (ASCII)	102,736 bytes	
		Trawl Data (ASCII)	116,289 bytes	
		Fish Tissue Chemistry (ASCII)	57,474 bytes	
		CTD Data (ASCII)	846,339 bytes	
		Toxicity (ASCII)	28,526 bytes	
		1990 Reference Survey (ZIP file)	16KB	
		1985 Reference Survey (ZIP file)	21 KB	
		1977 Reference Survey (ZIP file)	33KB	

C.4 Types, Volumes, and Status of Other Data

Resource Group	Study Type and Duration	Data Set Type and Format	Approximate Data Volumes	Repository and Status
National Surface Water Survey	Consists of three studies: Eastern Lake Study (ELS), Western Lake Study (WLS), and National Stream Survey. Quantification of the extent, location, and characteristics of sensitive and acidic lakes and streams in eastern and western U.S. Sampled during spring, summer, fall of 1984-1986.	All data sets in ASCII comma-delimited format with header records **ELS** Phase I (1984) Chemistry Survey Phase II (1986) Spring Seasonal Chemistry Summer Seasonal Chemistry Fall Seasonal Chemistry Summer Zooplankton Species/Abundance Lake Bathymetry Indices Summer Chlorophyll **WLS** Phase I (1985) Chemistry Survey **NSS-I** Pilot Study (1985) Chemistry Survey Field Site Observations Synthesized Data Phase I (1986) Chemistry Sampling	1798 records-1MB 146 records-107KB 270 records-235KB 239 records-186KB 491 records-178KB 129 records-9KB 221 records-24KB 752 records-597KB 339 records-208KB 1068 records-316KB 34 records-18KB 1765 records-1MB	

Appendix D
Preliminary Design and Options Document

D.1 Purpose
D.2 Option–Enhancement to EMAP Oracle Database to Handle Complex Data Types
D.2.1 Option Description
D.2.1.1 Data Directory and Data Catalog Enhancements
D.2.1.2 EMAP Public Web Site Enhancements
D.2.1.3 EMAP Internal Web Site Enhancements
D.2.2 Option Analysis
D.2.2.1 Online Resource Availability and Maintenance
D.2.2.2 System Configuration and Connectivity
D.2.2.3 Flexibility of Design to Adapt to Future Technological and Program Changes
D.2.2.4 User Satisfaction
D.2.2.5 Benefits and Costs
D.2.2.6 Risks and Contingencies
D.2.2.7 Recommendation for Implementation
D.3 Conclusions and Next Steps

D.1 Purpose

The existing EMAP–IM system is considered robust and flexible enough to be enhanced to meet current EMAP requirements. However, to have a system capable of supporting a national data integration program over an extended period, it is necessary to plan for options and vision of the possible future needs of the program. The current EMAP system only uses the Data Directory portion of the early EMAP Oracle Information Management System (U.S. EPA 1994b). As the program evolves, it may be desirable to add new functionality. The need for additional components will depend on factors such as the volume and variety of data collected by new field programs.

This document fulfills EEI-2 documentation requirements by reviewing future options for enhancing the EMAP–IM system as requirements evolve. The option is described and analyzed for feasibility, cost effectiveness, and the manual and automated functions needed for the option to work. The

options description is conceptual and does not specify a detailed system design. Rapid changes in information technology mean that detailed specifications should be delayed until implementation is imminent to ensure that the most effective hardware and software are utilized. EMAP should not limit its success by specifying these details prematurely.

D.2 Option–Enhancement to EMAP Oracle Database to Handle Complex Data Types

The current EMAP–IM system provides a useful index to the complex array of data types from research partners, but it is not configured for storing and disseminating actual data sets in a searchable database. If EMAP expands its role in storing and distributing EMAP data, it will need to enhance the system to enter data sets into a relational database so they can be queried by end users. This option explores development of a more sophisticated system to handle more complex data types, which include tabular data sets, maps and GIS coverages, remote sensing indicators models, complex electronic documents, and combinations of the above.

The option presented here would be an extension to the existing system. It would be implemented in response to EMAP program needs as they evolve with the EMAP research strategy and the CENR framework, which would generate a greater complexity and variety of data types, generators, and users. Since Working Group research plans are in their formative stages, additional requirements will arise and will be considered in future versions of this Plan. Further requirements from national and international standards and monitoring committees will also be addressed in the future.

D.2.1 Option Description

Under this option, the Data Directory would still form the core of the system and perform its current function of providing access to data. EMAP data would still be managed at distributed sites and made available via the Internet. However, the system would need to add a combination of data access techniques, such as:

- storage of data in relational databases with SQL access;
- storage of documents in native formats, (e.g., Microsoft Word, WordPerfect, other word processors);
- data in ASCII (e.g., CSV) format;
- map data stored in spatial databases for ad hoc queries; and
- maps stored as objects for downloading on request.

Implementation of this option could involve implementing an object-oriented data base accessible through Web server interfaces. Object-oriented technology would improve the system's capability

to handle the variety of data formats and links that EMAP researchers are generating. Object-oriented databases are currently widely used as optimal data management tools for heterogeneous data sets like those in EMAP. In this model, different types of electronic files in many formats would be treated as objects and tracked in the Data Directory. The object-oriented model would add functionality to the Directory to cross-reference related file types to one another and link different elements of complex documents (i.e., the text, graphics, maps, tables, images, video, and audio of a document could be managed as separate files but related together for user access). The Web server tools would allow the system to be queried directly by users.

These changes would require EMAP–IM to develop additional standards for data storage, naming conventions, and other standards. These requirements would increase the complexity of the system and therefore increase the difficulty in implementing and maintaining such a system. However, a system of this type would give greater flexibility and the ability to organize a variety of different types of information in a readily accessible and maintainable form.

D.2.1.1 Data Directory and Data Catalog Enhancements

Under this option, enhancements to the Data Directory and Data Catalog would include:

- documenting database structures;
- developing a comprehensive Data Dictionary; and
- making the Data Directory directly accessible to the Internet through web server tools to ensure access to a wider audience.

D.2.1.2 EMAP Public Web Site Enhancements

Under this option, the need for duplication of information between the EMAP Public Web Site and the EMAP Internal Web Site would be reduced. The tools required to maintain the EMAP Public Web Site would be similar to those currently being used to maintain the EMAP Internal Web Site. Specifically, data and metadata could reside in databases maintained on the Internal server and be accessible via Web interfaces. Interfaces on the public site could be used to generate SQL code which would then be fulfilled by the database engine on the internal site. Security for databases and data objects would be performed using the native tools of the data management system as well as the public access site. Releasing data to a wider audience would simply mean changing the level of security on a data file or a data record.

D.2.1.3 EMAP Internal Web Site Enhancements

For this option, the integration of Web server technology, HTML, and database structures should result in a system in which access to the databases can be controlled without isolating the internal web site and its useful information. The information which is currently distributed via the EMAP

Intranet would be available as password protected files, accessible through a number of levels of password protection.

This option would add the following products and functionality to the EMAP system:

- improved access on EMAP Public Web Site to Directory and Catalog;
- FTP site for transferring files among participants;
- a web site more integrated with the information system being used to organize and manage the data; and
- implementation of a map server (e.g., MapObjects, Maps On Demand) or spatial data engine for distributing geographic coverages.

D.2.2 Option Analysis

This section provides analysis of the described option according to EEI requirements.

D.2.2.1 Online Resource Availability and Maintenance

Maintenance for a system as described in this option is extensive, requiring a sophisticated staff, knowledgeable in the scientific information being disseminated. Within this scenario, the analysis and processing of the data are still done at individual programs and analysts' computers. This option would not include the generation, dissemination and maintenance of analysis tools. The expertise for interpreting and aggregating results would not be included in the system or its direct support staff.

The additional complexity of the information system would eliminate some of the version control problems in the current system. It is likely, however, that some flat files as currently used would remain in the system since they are the most effective way for distributing simple data files. Controlling access to a wide variety of resources within a complex system makes security management issues more difficult. Specifically, the accessibility to raw versus QA'd data would have to be controlled via security tools. The system would have to be configured to ensure that different levels of access are available to control the level of data accessibility that different users would experience. Access control would have to be implemented for each data set generated and for each level of processing and aggregation applied to a data set. This would be a significant undertaking and would require at least one dedicated staff member.

D.2.2.2 System Configuration and Connectivity

Using this option, data availability would be controlled by EMAP–IM (AED) through the EMAP Public Web Site. All data would be available for read-only access to anyone through a Web browser. EMAP–IM (AED) would be responsible for identifying major user categories and could control access to files and database entries using these user group categories. Primary data users, including

principal investigators from other organizations, could be given access to data by password. Secondary users would access a combination of databases and data objects including text files, complex documents, and flat files. The Data Directory would remain the main tool for locating data sets of interest. However, hyperlinks and text searches could also be implemented. Cross-references between the Data Directory, databases, data objects, and hypertext documents could be implemented using standard web site and database technology. Documentation would be cross-referenced with the Data Directory and also within the actual databases containing summary information.

The implementation of a Z39.50 compliant directory will help EMAP IM from two perspectives. First, the EMAP Data Directory will be made available to a wider audience as part of the Z39.50 international network of interoperable directories. Second, EMAP data would be more easily located by users outside EMAP who are performing Z39.50 queries and searches.

D.2.2.3 Flexibility of Design to Adapt to Future Technological and Program Changes

To ensure that the system can be expanded and enhanced, it is preferable to use industry-standard software, hardware and standards. Use of standard products like Oracle, web browsers, standard HTML, Z39.50, and other industry standards will make enhancements easier because the technology is readily available at low cost. This option presents a model for managing complex data that is also compatible with the information management technology used by most data clearinghouses (e.g., CIESIN) for effective data delivery. Implementation of this option according to such widely accepted models would ensure easy expansion of the system. As new capabilities emerge, it is likely that the expansion and enhancement of standard systems would be considered by vendors and developers.

D.2.2.4 User Satisfaction

This option would improve access to EMAP information for all users because it allows components maintained on the internal site to be served to the public access site without the current duplication and conversion of components, and isolation of the internal site. Data exchange and distribution would be improved as data access for users is simplified. The underlying system would be more complex and require more maintenance, but data access would be simplified because the Data Directory and the data would be co-located (logically). Increased standardization would simplify accessing and combining data from different programs. Readily available documentation would improve the probability of the data being used properly. Compatibility with Z39.50 would make a wider range of data accessible for users through searches and result in more potentially pertinent data being located with each search. This option would also eliminate the current deficiency in the system whereby partners who are not EPA employees are unable to access data under development.

D.2.2.5 Benefits and Costs

As indicated above, this option would improve and simplify user access to EMAP data. Additional costs to implement this option include personnel costs, hardware costs, and software costs. The object-oriented model, implemented with web server tools, will allow retrieval of a greater variety

of data types with greater specificity than is possible in the current EMAP–IM system. In addition, tools for delivering specialized data (e.g., GIS) can be incorporated, and improved searching and on-demand mapping could be added. Approximately two additional technical workers would be required over the current staffing levels. One person would act as the database administrator for the system, taking responsibility for security and consistency of the database designs. The second person would be responsible for interface development and support. Existing staff would still be required for system maintenance, data maintenance, data librarian, user support, and program management.

D.2.2.6 Risks and Contingencies

A system of this type would require significant commitment on the part of all participants to adhere to established standards. Database designs would have to be well established and followed in order for data to loaded and made available via the central system. Data could still be managed in a distributed mode, but this would require professional database administrators accessible for each site. If the data are managed in a distributed mode, there are risks that the entire database will not be searchable. In addition, network problems could result in unacceptable performance. Combining data from disparate designs, sampling plans, and analytical methods can be difficult. The data cannot be used effectively without extensive documentation. Contingencies could include centralization of all data sets that are pertinent to the EMAP database on a nightly basis or based on last time of update. These backups could be invoked in the case of major network failures. Backups of all data files could be made on all participating nodes interested in acting as backup sites.

D.2.2.7 Recommendation for Implementation

It is likely that the EMAP system will need to move to a more robust information system as the program continues to grow. The option discussed in this section could be implemented if the users find it difficult to access and effectively use data once they have been located using the Data Directory. This is likely to be the case when data with different formats are combined for complex analyses. Design of the database structure may need to consider segregating data by program and sampling plan, resulting in many discrete entities. Integrating the disparate data types will be difficult. Emphasis will be placed on linking the data with the appropriate metadata. Getting all the data to fit into a single structure is not the most important issue and is considered secondary to facilitating appropriate access by users.

D.3 Conclusions and Next Steps

This option will likely be implemented over the next few years. It is a logical next step and will keep EMAP current with emerging information management requirements and efforts to integrate EMAP into the larger environmental monitoring community.

EMAP will also begin exploring methods for facilitating access to other major environmental information systems including (but not limited to):

- EIMS;
- STORET;
- Envirofacts;
- NAQWA;
- ESDIM; and
- CIESIN.

EMAP will coordinate with other parallel efforts underway to develop data directories for monitoring data, including the EPA Environmental Information Management System (EIMS) directory.

Appendix E
Responses to "Environmental Monitoring and Assessment Program: Data Management Review Team Report"

E.1 Background
E.2 Review Team Comments and EMAP–IM (AED) Responses
E.3 Review Team Members

E.1 Background

The May 1996 version of the EMAP Information Management Plan was reviewed by an EMAP Data Management Review Team, who submitted their comments to EMAP–IM (AED). EMAP–IM (AED) revised the Plan in response to these comments to produce the October 1996 (U.S. EPA 1996a) version and inserted comments about the changes to the body of this review. Additional changes were made to the Plan to produce the current version, which incorporates the results of the Requirements Analysis interviews; further comments have been inserted into this review based on the new version. Responses to the reviewer comments are listed in two different typefaces, as follows:

Original Review Team comments (plain typeface)

EMAP–IM (AED) responses, December 1996 (italic typeface)

EMAP–IM (AED) RESPONSES, September 1998 (ITALIC TYPEFACE, ALL CAPS)

E.2 Review Team Comments and EMAP–IM (AED) Responses

We appreciate the time the Review Team spent reviewing the EMAP Information Management Plan and their recommendations that have resulted in an improved Information Management Plan.

The Information Management (IM) Plan has been revised to reflect comments from the Review Team and is in final form with regard to Phase I of EMAP IM. This Plan will continue to be updated as EMAP evolves. Certain aspects of EMAP IM for Phase II and Phase III are not yet completely defined. Finalization of the EMAP Research Plan (U.S. EPA 1996a) and the subsequent

implementation plans from the new EMAP Working Groups that identify the IM requirements of the revised EMAP will help refine the scope of the IM Plan. ORD Division representatives on the EMAP Information Management Working Group will assist in writing sections pertaining to those parts of the new program for which their Division have responsibility. The Plan will be forwarded to the ORD Senior Information Resources Management Officer and the ORD Information Management Strategic Plan group for review and comment.

We have responded to all of the Review Team's recommendations. Most of these recommendations are reflected in the revised Plan; a few recommendations pertain to information management practices of the entire Office of Research and Development and are more appropriately addressed in the new ORD Information Management Strategic Plan (U.S. EPA 1996c).

The following text in normal font is from the Nov 1, 1996" EMAP Data Management Review Team Report"; the text in italics is the AED/NHEERL response. References are made to sections of the revised IM Plan where the Review Team's comments are addressed.

Summary of Recommendations

The review committee decided to address the Environmental Monitoring and Assessment Program (EMAP) as three distinct phases of activity that are in different stages of development and whose information needs vary from known to uncertain. The committee believes that it is important to consider the current status of each the three phases in developing the information management plan. The uncertainties in the requirements for Phase III require a strong and defined program management plan. In general the committee feels that since Phase I is well understood, the plan should be revised to clarify and clearly define the data requirements, system structure and users, and begin entering the EMAP data from 1990 through 1995. Before Phases II and III of the information management plan can be implemented the committee believes that considerable effort must be made to define the database size, the users and their various needs, links and interfaces with other systems, and development of a clear management plan. The remainder of this report provides more detailed recommendations and discussions of actions to be taken in the areas of defining requirements, system architectures, administrative interfaces, and the development of the management plant.

Introduction

The redirection of the EMAP program from its original design as a national monitoring program to one focusing on developing a better scientific understanding of environmental measurements and their meaning in assessing ecological conditions, required a revision in the approach to information management. The EMAP Data Management Review Team met on May 22–23, 1996 to review the EMAP Information Management Plan: 1996–1998 prepared by the National Health and Environmental Effects Research Laboratory. The basic charge to the review team was "to determine if the information management approach being proposed to address the new direction of EMAP is consistent with and appropriate for addressing the information management needs of the research

program." The review team was to consider whether the goals and data management needs for EMAP have been defined and adequately described, are the system resources (e.g., hardware, software, personnel, etc.) sufficient, and are the organization and management roles and responsibilities as described adequate to support the task.

The revised EMAP is an interagency endeavor which is still evolving. Many uncertainties exist regarding all the active participants (e.g., Federal agencies, EPA organizations, extramural researchers) in the program, what data will be generated and by whom, and who all the ultimate users of the data might be. These uncertainties make it very difficult to design the information management system and presented difficulties in evaluating the overall approach. The review team decided to address the EMAP program as three distinct phases that are in different stages of development and whose information management needs vary from known to uncertain. Phase I focuses on storing and managing the data collected in the original EMAP program between 1990 and 1995. This component of the program is known and can be very well defined. Phase II involves the handling of data currently being collected at Index Sites and regional studies such as the Mid-Atlantic Integrated Assessment (MAIA) study. Phase III of the EMAP (EMAP2) is still being planned as an interagency effort through the Committees on the Environment and Natural Resources (CENR). Uncertainties such as the specific data to be collected, entered and maintained in any EMAP2 information management system add to the difficulty in designing the ultimate data management system.

Just to clarify: For Phase I (the EMAP 1990–1995 data), we include the EMAP Data Directory and Catalog and the EMAP home page activities that are an integral part of managing and distributing those data. For Phase II, we add data management activities from all of the new EMAP Working Groups, which include—in addition to index sites and regional studies—ecological indicator research, multi-tier design, and REMAP. Most of the Phase II studies will be pilot projects or integral parts of the CENR (Phase III).

The review committee believes that it is important to consider the current status of each the three Phases in developing the information management plan, and that the uncertainties in the requirements as EMAP2 develops requires a strong and defined program management plan.

The IM Plan has been revised to reflect the three Phases suggested by the Review Team. More detail has been added (Section 5, Project Management and Coordination) about the current status of each of the three phases.

[IN THE FINAL PLAN, CENR IS NOT REFERRED TO AS PHASE III BECAUSE IT IS AN ONGOING EFFORT THAT IS PART OF PHASE II AND FUTURE EMAP PLANNING. ISSUES OF CENR INVOLVEMENT ARE CLEARLY EXPLAINED IN THE PLAN AND HANDLED AS AN ONGOING ISSUE IN SUCH SECTIONS AS: APPROACH (2.4), USERS (4.2.2); DATA SOURCE TRACKING REQUIREMENTS (4.4.1); DATA STANDARDS (4.4.3); DATA DIRECTORY, DATA

CATALOG, AND WEB SITE FUNCTIONALITY (5.4.1, 5.4.2, 5.4.3); PROGRAM MANAGEMENT (6.7.2, 6.7.3), AND IMPLEMENTATION (7.1.3, 7.1.5, 7.2.5, 7.2.7, 7.4, 7.5).]

The EMAP Research Plan (U.S. EPA 1997a) has been revised and will be supplemented with more detailed EMAP research implementation plans with input from each of the new EMAP Working Groups. These implementation plans for each new Working Group should address descriptions of new data types, volumes, and collection schedules. The IM plan was revised based on the Review Team recommendations; it will be updated after the EMAP Research Plan and Working Group implementation plans are completed (as details of Phase II and Phase III become available).

[INFORMATION ABOUT THE NEW DATA TYPES AND NEEDS OF EACH WORKING GROUP HAVE BEEN COLLECTED FROM WORKING GROUP IMPLEMENTATION PLANS AND REQUIREMENTS ANALYSIS INTERVIEWS AND ARE NOW INCLUDED IN SECTIONS 3, 4, AND 6.]

Development of a program management plan will be discussed with EMAP managers. The EMAP Director position has now been filled, and this issue will be discussed with him.

[A PROGRAM MANAGEMENT PLAN IS PRESENTED IN SECTIONS 6 AND 7 OF THE PLAN THAT DISCUSSES THE ORGANIZATIONAL AND RESOURCE ISSUES RELATED TO IMPLEMENTATION OF THE PLAN.]

In general, the committee recommends that since Phase I is well-understood, the plan should be revised to clarify and clearly define the data requirements, system structure, and users and begin entering the EMAP data from 1990 through 1995.

Phase I requirements, system structure, and users have been clarified and defined in the Plan (Sections 2.1, 6.3).

We are focusing on the known requirements of Phase I: 1990–1995 data; data distribution via the EMAP home page, and the Data Directory. The process of loading EMAP 1990–1995 data sets and metadata files to the EMAP home page is well under way. A status report on availability of these data has been included in the Plan (APPENDIX C, INVENTORY OF DATA TYPES).

Before Phases II and III of the information management plan can be implemented the committee believes that considerable effort must be made to define the database size, the users and their various needs, links and interfaces with other systems, and development of a clear management plan.

The Information Management Working Group (IMWG) will coordinate with other EMAP Working Groups to help them define their IM needs as their program plans evolve. AED will use the Information Technology Architecture Support (ITAS) contract to help scope out Phase II and III

needs and requirements and to incorporate this information into an update of the Plan, targeted for 1997B8. This will include information on data types, volumes, and collection schedules for Phases II and III.

[PHASE I AND II NEEDS HAVE BEEN DEFINED IN REQUIREMENTS ANALYSIS INTERVIEWS. THEY ARE PRESENTED IN SECTIONS 3 (DATA) AND 4 (REQUIREMENTS), AND APPENDIX A (NEEDS AND REQUIREMENTS DOCUMENT.]

The IMWG includes members who are also on the CENR Task Force on Data Management; these people will help identify CENR IM requirements that EMAP IM could address.

More information on program management has been added to the Plan (Section 6). Recommendations regarding program management have been brought to the attention of the EMAP managers.

The remainder of this report provides more detailed recommendations and discussions of actions to be taken in the areas of defining requirements, system architectures, administrative interfaces and the development of the management plan.

Requirements

Design of an Information Management System (IMS) requires knowledge of the types and amounts of data to be stored and the ease of accessibility intended for primary and secondary users. The IMS Plan presently lacks sufficient specificity for practical guidance. In terms of Phase I the needs are undoubtedly known to EMAP, but are not specified in the IMS plan. The Phase II needs may be relatively simple but should be documented. The Phase III needs are vague because no one knows the EMAP responsibilities under the CENR-inspired monitoring of Index sites.

More detail has been added to the plan about the types and amount of data in Phase I [APPENDIX C, INVENTORY OF DATA TYPES]. <u>The EMAP Research Plan</u> to be used for Phases II and III is not detailed enough to determine whether the simple approach planned for Phase I—a Data Directory, and data sets and metadata files accessible through the EMAP home page—is adequate or whether the Phase II and III data sets to be managed are of sufficient size and complexity that a more complex Oracle database management system would be required.

[THE BEST AVAILABLE INFORMATION ABOUT PHASE II DATA TYPES, VOLUMES, AND STATUS ARE SUMMARIZED IN APPENDIX C, INVENTORY OF DATA TYPES. OPTIONS FOR EXPANDING THE CAPACITY OF THE EMAP IM SYSTEM TO MANAGE COMPLEX DATA TYPES ARE EXPLORED IN APPENDIX D, PRELIMINARY DESIGN AND OPTIONS ANALYSIS.] PHASE II IS STILL IN A DESIGN PHASE AND THE VOLUMES OF DATA TO BE MANAGED, AND THE ANALYSES THAT WILL BE PERFORMED ON THE DATA ARE STILL NOT COMPLETELY DEFINED. SOME OF THESE UNKNOWNS WILL NOT BE CLEARLY UNDERSTOOD UNTIL SOME TIME IN FY 1999.

Phase I

The objective of Phase I is to get all the historic EMAP-collected data of 1990–1995 into a single and readily accessible database. The first order of business is not IMS design but simply getting all the data into the existing database.

The objective of Phase I is to make the 1990–1995 data publicly available on the Internet (EMAP WWW site). A database to put all the data into—in the sense of a database managed by database management system software such as Oracle—does not exist. This has been clarified in the revised Plan. The process of loading 1990–1995 EMAP data to the EMAP home page is well underway.

The committee probably has little idea of how much data EMAP does.

Recommendation: The IMS plan should contain listings of the datasets that constitute the EMAP database, their approximate sizes when full, and estimates of how much data still need to be added.

An inventory and a status report on 1990–1995 data sets are being updated as the information is gathered from the old Resource Groups; these reports have been included in the Plan [Appendix C, Inventory of EMAP Data] .

Since the primary users, EMAP scientists, have written reports based on stored data, it is obvious that they can extract data. It is evident from the EPA Region III Mid-Atlantic Integrated Assessment (MAIA) that at least one secondary user has extracted data. It was stated in the plan that EMAP has shed a prior IMS feature that provided users with statistical tools.

Recommendation: Further simplification of the system may be in order. EMAP is encouraged to solicit comments from potential secondary users on their success in getting the data they sought and difficulties encountered.

Providing access to data sets and metadata files via the EMAP public access World Wide Web home page is quite straightforward; it is not anticipated that it could be simplified much further. However, users of the system will be solicited for their suggestions for improvements.

Once the data are available on the EMAP Web Site, the problems of secondary users accessing data should be solved. The Web Site includes a comment section where users can respond with questions or comments. The comment later in this report regarding user advisory groups will be discussed with the EMAP managers.

Phase II

The IMS is ambiguous on whether the data base will manage data that were not generated by EMAP-funded projects. The plan recognizes the environmental assessments such as the MAIA use more than just EMAP data. The plan needs to be clear, however, that EMAP will provide a directory of monitoring datasets compiled by other federal, state, and local agencies but will not provide the raw data. Users on the Internet will be linked to those other datasets.

Recommendation: It should be clearly stated in the IMS plan that EMAP does not intend to manage data that were not collected with EMAP funds.

This has been clarified in the Plan ([SECTION 2.5.2]).

The IMS is not clear on the detail that EMAP intends to provide for these non-EMAP data bases. Since this could range from the complete metadata for each data base to only a data set name, there is a need for specificity here. It appears that EMAP intends to provide the metadata for all the data bases for which it provides link. Since the metadata is already residing with the raw data in the host server, this would be an unnecessary redundancy.

METADATA REQUIREMENTS ARE BEING SPECIFIED BY BOTH EMAP AND THE GENERAL USER COMMUNITY. MINIMAL METADATA WILL BE REQUIRED FOR ALL DATA SETS. UNLESS MINIMAL DOCUMENTATION is AVAILABLE FOR A DATA SET, IT WILL NOT BE MADE AVAILABLE THROUGH THE EMAP IM SYSTEM. IN SOME CASES, THE EMAP IM TEAM WILL DEVELOP THE MINIMAL DOCUMENTATION FOR ORPHAN DATA SETS.

Recommendation: The plan should specify how much detail EMAP intends to provide for non-EMAP data bases, and should evaluate the need to include metadata contained in the linked data base.

RESOLVED THROUGH RECOMMENDATION OF USING THE FGDC STANDARD, WHICH INCLUDES MINIMAL REQUIREMENTS.

Only Directory information on other databases will be provided. Table 5-2 has been added to the Plan that lists the Data Directory fields. EMAP does not intend to provide metadata for the databases for which the EMAP home page provides a link. This has been clarified in the IM Plan [SECTION 2.5.2, EMAP-IM APPROACH]. Table 5-1 shows what EMAP-IM features (Directory, Catalog, data sets, links) are to be used with different sources of data has been added to the Plan.

The IMS contains wording to the effect in the plan that EMAP will provide instructions to other monitoring programs so that their data meet certain standards. This could be a mammoth task that breeds only resentment and eventual disregard among the "guided" agencies.

Recommendation: EMAP should make clear statements about what data standards it intends to recommend for non-EMAP databases.

EMAP will provide guidance to EMAP-funded programs, not to non-EMAP monitoring programs. Guidance to the latter is coming from the multi-agency CENR monitoring framework. This has been clarified in the Plan [SECTION 4.4.3].

WHEREVER APPLICABLE NATIONAL AND INTERNATIONAL STANDARDS WILL BE ADOPTED. THE EMAP WEB SITE WILL CONTAIN LINKS TO DESCRIPTIONS OF THE ADOPTED STANDARDS. IN SITUATIONS WHERE CENR HAS ADOPTED STANDARDS, EMAP WILL ADHERE TO THESE STANDARDS AND WILL ALSO TRY TO BE COMPATIBLE WITH OTHER REGULARLY USED STANDARDS.

Phase III

The requirements for data to be collected in 1996 and thereafter are unknown and may remain so for some time. There are three sources of new data; the Regional EMAP programs, the principle investigator-initiated research to be funded by EMAP, and data gathered at Index sites as defined by the CENR. It would seem that EMAP should manage whatever data are collected under its sponsorship. However, not all REMAP data are in that category, not all PI data will be applicable to environmental assessments, and many agencies aside from EPA will be gathering data at Index sites.

Recommendation: The data management requirements for Index Sites may defy definition at this point but some attempt could be made to delimit the other Phase III needs.

The revised EMAP Research Plan, with implementation plans for the new EMAP Working Groups, should help clarify this. The EMAP IMWG intends to meet with the other EMAP Working Groups and to use the ITAS contract support to help define the IM needs of Phase III.

THE NEW PLAN CONTAINS INFORMATION ABOUT NEEDS OF PHASE II AND BEYOND (SECTION 4), AND EXPLORES FUTURE OPTIONS FOR MEETING THESE NEEDS AND OTHER POTENTIAL NEEDS THAT RESULT FROM PROGRAM EXPANSION (APPENDIX D, PRELIMINARY DESIGN AND OPTIONS ANALYSIS)

Architecture

It is essential to view and understand the current architecture for EPA's existing systems that may provide a base for further development. A number of specific actions should be taken to enhance the system's architecture:

Recommendations:

Any reusable "modules" of the system should be ported into the redesigned system.

More information has been added on the Oracle development work done prior to 1996 and how it is being used (Section 5.2).

Populated databases should also be ported to updated/upgraded database structures.

Data from the Data Directory database were ported over to the revised Data Directory (Section 5.4.1).

The network configuration, the software and hardware configuration (including future host machines and operating systems), socket level functionality, user interfaces, database access, and resulting data flows must be clearly documented and evaluated for migration, integration, implementation, and extensibility.

More information has been included on the existing EPA computing infrastructure—computers, software, networks, personnel (App. C—THIS INFORMATION IS NOW IN SECTION 5, TECHNICAL DESIGN).

Planning for external technological innovations should also be considered.

EMAP, through the Laboratory Divisions and with support from the Enterprise Technology Services Division of OIRM, 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 (Section 4.4.4). We will apply these to the EMAP system when practical.

The resources necessary for accurate system documentation should be established and acquired.

EMAP has used Oracle CASE tools for system documentation.

No structural diagrams or words describing thoughts and/or plans for distributed architecture were apparent. Gateway access among computer systems must be carefully designed. Web links do not constitute a distributed system.

More detail has been added (Section 5, FIGURE 5-3). Links to other Web sites are not what was meant by distributed. What was intended to be described was the existing database management systems at the various EPA Labs; this has been clarified (Section 2.5.2).

Economies of scale and relationships among metadata components must be streamlined. Draw directly from data to populate inventory, catalog, and directory, or minimally draw directly from the data to populate the inventory; draw directly from the inventory to populate the catalog; and draw directly from the catalog to populate the directory.

We explored development of a combined Data Directory/Data Catalog database. The Data Catalog is the metadata that the user needs in order to use the data wisely. Database design for the Catalog was done by the former central EMAP–IM group; however, this was never implemented in Oracle. Instead, the program used ASCII text files; some of these were loaded to Oracle Book, which is a hypertext-based product. The advantage of a combined Data Directory/Data Catalog database is that common fields—such as the contact telephone number—need to be modified in only one place. However, the Catalog is mostly text-based, so this advantage does not often apply. Also, we expect to have far more Data Directory entries for data sets than we expect to have for actual data sets in our possession (thereby requiring the Catalog information). It is much easier to acquire Directory information than it is to acquire the full set of Catalog information.

Our plan is to keep the Catalog separate from the Directory and to produce the Catalog in ASCII text (a WordPerfect template is used) and then mark the text up in the HTML format. As we are not expecting our users to have access to Oracle Book, but rather to use Web browsers, this should be a practical approach. The EPA National Center for Environmental Assessment (NCEA) is currently exploring an Oracle implementation of the Data Catalog. This will be evaluated for potential application to EMAP. We are also keeping track of what the NOAA Coastal Services Center in Charleston has done to extend the FGCD metadata standard to field sampling data, and the use of metadata entry tools (NOAA 1996). Future updates to the EMAP IM Plan will reflect any changes.

AT THE TIME OF THIS WRITING, MANY TOOLS HAVE EMERGED TO STANDARDIZE THE DEVELOPMENT OF METADATA. IN ADDITION, THE FGDC STANDARD HAS EMERGED AS A CORE SET OF DOCUMENTATION ELEMENTS WHICH CAN BE USED AS A POINT OF DEPARTURE FOR THE DEVELOPMENT OF COMPREHENSIVE DOCUMENTATION. FGDC WAS ORIGINALLY DEVELOPED TO DOCUMENT SPATIAL DATA COVERAGES BUT IS BEING EXPANDED TO DEAL WITH OTHER DATA TYPES. AS THIS STANDARD AND THE TOOLS AVAILABLE TO USE IT CONTINUE TO GROW, EMAP AND MOST OTHER PROGRAMS WILL ADOPT THE STANDARD AS THEIR DATA CATALOG FORMAT. FGDC IS SET UP AS A HYPERTEXT FILE WITH TAGS THAT CAN BE USED EFFECTIVELY WITH HTML OR SGML. IN ADDITION, HYPERTEXT WILL BE VERY COMPATIBLE WITH EXTRACTING SOME INFORMATION AND LOADING IT INTO A DIRECTORY OR INVENTORY DATA BASE. THIS IS POSSIBLE BUT IS NOT AS STRAIGHTFORWARD AS RECOMMENDED BY THE REVIEW TEAM ESPECIALLY IN REFERENCE TO THE INVENTORY, IN WHICH SOME DESCRIPTION OF THE PROGRAM IS REQUIRED THAT IS NOT LIKELY TO BE INCLUDED IN THE DOCUMENTATION. IN ADDITION, MANY PROGRAMS OVERLAP AND DIFFERENT NAMES ARE USED FOR THE SAME PROGRAM (ESPECIALLY AS THEY CHANGE THEIR NAMES OVER THE YEARS).

No database schema was included in the IMS. A data base schema and data dictionary constitute the foundation of a stable information base that will permit portability and access by an expanding and

demanding user population. Decisions on maintaining the metadata and/or the actual data in the database are critical in the schema.

The Entity-Relationship Diagram (Fig. 5-4) and associated Table 5-2 for the Oracle Data Directory have been included in the Plan.

The IMS should clearly state the standards with which the EPA wishes to comply. Choosing standards that permit compatibility with other search and retrieval or analysis packages could provide considerable savings. Also, choose flexible, interoperable standards that allow evolution with changing demands. Example: Choose an "attribute set" standard used by Z39.50. As the Z39.50 upgrades to include numeric and geographical searches, the information collected and held in the appropriate format (fields) can permit immediate search and retrieval.

EMAP will follow applicable EPA and federal IM standards. EMAP is using the EPA public access World Wide Web server at Research Triangle Park, NC, as the site where users will do searches, and will encourage RTP to make their. WAIS (Wide-Area Information Server) software Z39.50 compliant. EMAP has adopted the restricted vocabulary from the Global Change Master Directory to be used in the EMAP Data Directory (Section 5.4.1, 5.4.3,).

Develop partnerships with other Federal agencies that are performing similar work. Cooperative agreements and good planning can permit the integration of technology developed elsewhere without the need to duplicate effort.

We have contacted NASA about the Global Change Master Directory, use of a restricted vocabulary keyword set, FGDC compliance, and use of DIF entry tools on the World Wide Web. This is reflected in the Plan (Section 5.4.1).

We will work with the CENR Task Force on Data Management, via the EMAP Information Management Working Group members on the Task Force, to coordinate with other agencies (Section 6).

"Maintenance" should be considered when developing an architecture. Developing reusable tools can expedite the completion of work.

Tools for making Data Directory entries are being revised/developed. These include an Oracle Forms 4.5 client application and a Web form to capture directory information that is later loaded to the Oracle database (Section 4.3).

[THE EMAP IM SYSTEM COMPONENTS REQUIRE SIGNIFICANT ONGOING COMMITMENT OF HUMAN RESOURCES TO MAINTAIN CONTENT, AND PERIODIC UPGRADES TO HARDWARE, SOFTWARE, AND INFRASTRUCTURE TO TAKE ADVANTAGE OF

TECHNOLOGICAL IMPROVEMENTS. THIS INFORMATION IS REVIEWED IN SECTIONS 6 (PROJECT MANAGEMENT) AND 7 (IMPLEMENTATION).

Administrative Interface

Review against Office of Research and Development (ORD), Agency and Federal requirements

The ORD Five-Year Information Resources Management (IRM) plan, the Agency Five-year IRM Strategic Plan, numerous Agency IRM policies and procedures, currently catalogued as EPADOC by the Architectural Management and Planning Branch, as well as Federal Information Management Regulations (FIRMR, FAR) are in place to provide a road map as well as an architectural framework for designing new or improving current information systems in the Agency. The review shows attention paid to the public access and public information policies proscribed in the above mentioned documents, but the life cycle methodology and data management system rigor necessary is not evident.

These requirements were addressed by the peer-reviewed 1994 EMAP Information Management Strategic Plan (Shepanek 1994) and earlier life cycle documents (e.g., U.S. EPA 1991). The current Plan does not discuss any new system elements that were not included in the old EMAP IM system and covered by the previous documents. AED will use the ITAS contract support to address life cycle requirements for data from the new EMAP Working Groups. (SEE APPENDIX H, EEI REQUIREMENTS REPORT)

The ORD Senior Information Resource Management Officer has appointed a representative from ORMA (Office of Resources Management and Administration) to the EMAP Information Management Working Group (IMWG). This person will be the liaison between EMAP and ORMA, OIRM, and other federal information management standards. This person can help EMAP stay informed of the latest changes in federal information management standards and could help update the EMAP IM Plan to meet those standards.

The ORD Information Management Strategic Plan (U.S. EPA 1996c), currently in draft stage, will be a key document. Updates to the EMAP IM Plan will show how EMAP IM fits under the umbrella of the ORD IM Strategic Plan as it evolves.

Two of the Agency's eight IRM vision elements of public access and EPA access have been addressed in part; however, the vision elements of data integration, environmental information, a solid IRM foundation, reduction of reporting burden, electronic management, and communications have not been addressed fully.

Recommendation: The IMS should address the six vision elements cited above.

We will ask ORMA to review and help update the IM Plan after the EMAP Research Plan is finalized. Specifically, the ORMA representative on the EMAP IMWG could assist in addressing all eight of these elements in the next update of the Plan.

ALL EIGHT ELEMENTS HAVE BEEN ADDRESSED IN APPENDIX I OF THE PLAN.

Use of Agency Resources

A wide variety of resources are available within the Agency to organizations contemplating a new information system or the redesign of one that exists. These resources reside primarily within the OIRM) and its counterpart office within ORD, the Management and Information Systems Staff (MISS). The Senior Information Resource Management Officer (SIRMO) has dual responsibilities as the MISS staff chief providing the necessary review and integration function for information systems throughout ORD, and considering all necessary information system interfaces with the Agency as a whole. Economies of scale must be leveraged as human and financial resources shrink, and the SIRMO provides the system integration and review function necessary to new and renewed information systems endeavors.

No evidence exists that the renewed EMAP system was reviewed by the ORD SIRMO, nor that these services were requested from the SIRMO. These review services are the mission of the SIRMO and are a critical part of insuring user group representation, particularly secondary users in Program and Regional Offices. In particular, life cycle methodology corresponding to a multi-year budget, multi-year schedule as well as integrating this system with current systems is lacking in the IMS. The SIRMO provides the necessary coordination to insure that infrastructure requirements are met, and cost/benefit analyses have been performed to the extent necessary to field such a system.

Recommendation: The revised IMS should be submitted to the SIRMO for review. The EMAP system as it stands today should receive no additional funding for further development until a complete and thorough SIRMO review has been undertaken.

The May 14, 1996 draft of the EMAP IM Plan was not submitted to the ORD SIRMO because of its incomplete state. It was incomplete because the EMAP Research Plan (U.S. EPA 1996a), and EMAP Working Group implementation plans, were not sufficiently detailed to be able to specify the new EMAP IM requirements. The May 14 draft was prepared to see if the general approach to EMAP IM was acceptable. After the EMAP Working Group implementation plans are completed, and the ITAS contract support has helped define EMAP Phases II and III IM needs, we will ask ORMA to review and help update the Plan.

Management

The plan lacks adequate descriptions of the management infrastructure to implement the EMAP IMS. Almost all problems encountered during development and operation, such as cost over-runs,

schedule delays, technical glitches, and inadequate performance, stem directly from poorly defined management organization or processes. A solid, well-organized management infrastructure has the capability to detect and correct minor problems before they balloon into major program issues. The management infrastructure maintains the focus and discipline necessary to keep the EMAP IMS on track and on schedule. The Implementation Plan must lay out the EMAP IMS management infrastructure as well as the technical implementation.

EMAP uses a matrix management approach implemented through Division line management. Project management is an issue that is common to many other ORD information management systems and is being addressed at a higher level by the Information Management component of the ORD Strategic Plan (U.S. EPA 1996c).

Recommendation: The committee suggests adding the following items to the Implementation Plan:

Organization Hierarchy—The Implementation Plan should contain the organizational hierarchy for the personnel in EMAP IMS. The hierarchy should emphasize the official reporting structure—who works for whom. In this context, reporting structure refers to who pays a person's salary and signs his performance evaluation.

This has been clarified in the revised Plan (Section 7).

THE IM SYSTEM IS ONLY WHAT IS BEING DEVELOPED AT AED THROUGH THE EMAP INFORMATION MANAGEMENT WORKING GROUP.

Functional Breakdown—The Implementation Plan should contain a mapping of personnel to functions, deliverables, and requirements. Essentially, this clearly defines development responsibility, eliminating the possibility of duplication or gaps.

This will be shown for Phase I in the next update to the Plan; other Phases will be added later (SECTION 7).

Multi-year Schedule—The Implementation Plan should contain a multi-year schedule clearly identifying EMAP IMS deliverables. The schedule should include the name and length of those development tasks that define the critical path. The schedule should also clearly identify the number of days of "slack" between the critical path and the product delivery dates. The schedule should also identify regular, formal design reviews associated with each product deliverable. The schedule should cover a five-year period, updated at the start of each fiscal year.

The revised Plan has more details on the schedule (Section7). However, details of the nature called for by the recommendation will not be available until the EMAP Research Plan, and the implementation plans supporting that plan—along with their multi-year schedules—are finalized.

Formal design reviews for each deliverable may be more than is necessary for the EMAP IM system, which is envisioned to be rather simple, consisting largely of data sets and metadata file accessible through the EMAP home page.

Multi-year Budget—The Implementation Plan should contain a multi-year budget. The budget should clearly identify the funds required for each system deliverable. The budget should clearly identify the funds required for normal operations, such as data ingest, maintenance, data delivery, user support, etc. The budget should also clearly identify the programmatic cash reserve to handle contingencies. The budget should cover a five-year period, updated at the start of each fiscal year.

The revised Plan has more details on the budget (Section 7). However, details of the nature called for above will not be available until the EMAP Research Plan, and the implementation plans supporting that plan, are finalized.

Management Processes—The Implementation Plan should contain descriptions of the basic processes to explain how the system works. For example, handing someone a copy of the Constitution would not help him to get a new law passed. The process descriptions should include basic configuration and version control of data content, software, and documentation. The process descriptions should include identification, allocation, and control of critical resources, in this case, probably data storage and processing. The process descriptions should include acceptance criteria and procedures for standards selection, design review, and design implementation. The process descriptions should also cover writing of documentation, and training of personnel and users. The process descriptions should identify and define any required control boards and advisory groups.

IT CAN BE DONE BY RESTRICTING THE "SYSTEM" TO WHAT EMAP–IM (AED) WILL DO.

The Plan has been revised (Section 6), but the management process discussed in the recommendation is probably more detailed than would be required for EMAP IM.

Interface Listing—The Implementation Plan should contain a listing of interfaces between the EMAP IMS and other organizations, both internal and external to EPA. The interface listing should clearly identify functional (system-to-system) and communication (group-to-group) interfaces. The Implementation Plan should clearly define the interface boundary and what gets transferred into or out of the EMAP IMS. The EMAP IMS should avoid duplicating the management processes when listing communication interfaces. Each interface should apply to one, and only one, organization or system external to the EMAP IMS. Detailed descriptions of the interfaces belong in separate documents and not the Implementation Plan.

As these interfaces become more clearly defined, from the new EMAP Working Groups or from activities of the CENR, we will include descriptions in the IM Plan.

User Group Representation—The Implementation Plan should contain a listing of all user advisory groups. The Implementation Plan should define the role, purpose, and product for each advisory group. The Implementation Plan should clearly identify when these groups meet in the master schedule. The Implementation Plan should identify the general membership of each advisory group (but not specific names). The Implementation Plan should map the advisory groups to the target user groups and the management processes.

At this time, EMAP does not have any specific user advisory groups. There were such groups in the early EMAP, consisting of personnel from EPA program offices, EPA regional offices, other agencies, and academia. This recommendation will be discussed with the EMAP managers.

Management Issues

The EMAP IMS must also face several serious management issues:

Work Force Fragmentation—The EMAP IMS has too many people working too little time each. The Implementation Plan identifies several individuals who devote as little as 10% of their time (0.1 FTE) to the EMAP IMS—only 4 hours per week. These individuals would barely have enough time to attend a meeting and make a few phone calls, let alone to do any serious system development. Also, these individuals will constantly get "pulled off" onto "higher priority" projects, further reducing EMAP IMS development time. Lastly, the time devoted to integration increases exponentially with the number of people on a project. Individuals that devote only a small portion of their time to a project then become a burden, not an asset.

Recommendation: The EMAP IMS should have fewer people devoting more of their time to system development. No individual should devote less than half of his time (50% or 0.5 FTE) to EMAP IMS development.

The recommendation will be discussed with EMAP managers.

Systems Integration—The EMAP IMS lacks people devoted to integrating the pieces into a single system and making it work. The different components under development by separate organizations (internal and external to EPA) will not magically fit together into a unified system.

Recommendation: The EMAP IMS should assign a single person the responsibility of systems integration. Systems integration is not a part-time job for a system of this size. This person must devote his full time (100% or 1 FTE) to the task of systems integration.

Computer science support positions are provided to ORD facilities through on-site ADP contractors. The recommendation will be discussed with EMAP managers.

Leadership—The EMAP IMS does not have a clearly identified leader. Without clear and strong leadership, the program will fail.

Recommendation: The EMAP Program should assign a person as Project Manager for the EMAP IMS. Project Manager is not a part-time job for a system of this size. The Project Manager must devote his full time (100% or 1 FTE).

The recommendation will be discussed with EMAP managers.

Authority and Responsibility—The EMAP IMS Project Manager has responsibility to implement the program, but does not have clear authority over budget, schedule, and personnel. Without this authority, the program will fail.

Recommendation: Place authority for personnel, budget, schedule, and technical content directly under the EMAP IMS Project Manager. Authority over personnel means the Project Manager signs yearly performance evaluations. Authority over budget means the project manager signs the paperwork controlling who gets what funds and when. Control over schedule and content means the Project Manager signs the paperwork setting requirements, passing design reviews, and approving product milestones.

The Chair of the Information Management Working Group does not have any supervisory control over any members of the Group. A major portion of the IM work for EMAP is conducted by personnel on EPA contracts (different contracts at the various Divisions). It would be illegal for these contract personnel to be under supervisory control of the EPA EMAP IMS Project Manager. The recommendation will be discussed with EMAP managers.

Budget and Resources—The Implementation Plan completely lacks any justification for the EMAP IMS budget, even though the EPA spends $1 million per year on it. The money devoted to system support and maintenance ($100k) appears extraordinarily high. The apparent data volume and ingest rate do not justify purchasing $130k in new hardware.

THIS IS CONTRADICTORY TO POINTS RAISED ABOVE ABOUT PERSONNEL, COMPLEXITY AND SIZE OF SYSTEM.

Recommendation: The Implementation Plan should contain justification for each line item in the EMAP IMS budget. The EMAP IMS should reduce the system support and maintenance to about $10-20 k per year. The EMAP IMS should defer any new hardware purchases until Fiscal Year 1997, at the earliest.

No hardware purchases were made in 1996. The budget, and justification, are more clearly delineated in the revised Plan (Section 6).

References

NOAA. 1996. Metadata resources fact sheet. NOAA Coastal Services Center, Charleston, SC.

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

U.S. EPA. 1991. EEI-1 Mission Needs Statement EMAP. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Washington, DC, July 1991.

U.S. EPA. 1996a. (in prep). EMAP Research Plan (July 1996 draft). U.S. Environmental Protection Agency, ORD, NHEERL, Research Triangle Park, NC.

U.S. EPA. 1996b. (in prep.). ORD Information Management Strategic Plan. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

E.3 Review Team Members

Name	Institution
David Kleffman	US EPA, ORD, National Center for Environmental Research and Quality Assurance, Washington, DC
Nancy Wentworth	US EPA, ORD, National Center for Environmental Research and Quality Assurance, Washington, DC
Robert King	US EPA, Office of Water/STORET, Washington, DC
Charissa Smith	US EPA, ORD, Office of Resources Management and Administration, Washington, DC
Dwight Clay	US EPA, Office of Information Resources Management, Research Triangle Park, NC
Michael Slimak	US EPA, ORD, National Center for Environmental Assessment Washington, DC
Lola Olsen	National Aeronautical & Space Administration, Greenbelt, MD
Kevin Schaefer	National Aeronautical & Space Administration, Washington, DC
Thomas O'Connor	National Oceanic and Atmospheric Administration, N/ORCA, Washington, DC
John Briggs	US Geological Survey, Reston, VA
Jim Slack	US Geological Survey, Reston, VA

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

F.1 Introduction
F.2 Data Sharing
F.3 EMAP Public Web Site
F.4 EMAP Data Directory
F.5 EMAP Data Catalog
F.6 Further Information

F.1 Introduction

This document briefly describes the information management policies, guidelines, and standards of EPA's Environmental Monitoring and Assessment Program (EMAP). Many of the EMAP documents referenced can be found on the EMAP Public Web Site (EMAP 1998). The EMAP Policy Statement is given in Section 1.

Research conducted by EMAP is increasingly dependent on data sets that are collected and managed by other organizations. A major challenge in conducting ecological assessments is the acquisition and integration of data of varying ownership, format, quality, and degree of documentation (ESA, 1995; NRC, 1995a). The difficulty is reduced as data sources move toward common standards, data directories, and data descriptions (Chinn and Bledsoe, 1997; Williams, 1997; Barton, 1996, 1997; LTER, 1995; CENR, 1994; FGDC, 1994). Many useful guidelines and standards have evolved from the U.S. Global Change Research Program (GCRP, 1995a, 1995b). These developments allow researchers to more easily find, understand, and download data of interest. The increasing use of World Wide Web browsers as a common user interface to distributed databases, that use a variety of software packages and database structures, also helps.

There are three basic components to the EMAP information management system: a Data Directory, a Data Catalog (metadata), and the actual data sets (U.S. EPA 1996a). The Data Directory keeps track of all data sets of interest. Some of these data sets are in the possession of EMAP and are on EMAP web sites; other data sets are managed and described by other organizations. The different data sources handle the data using a variety of software and hardware. The Data Catalog contains

information (metadata) about data sets in the possession of EMAP so that a user can understand enough about the methods, assumptions, and quality to use the data appropriately. With a Web browser, users locate data sets of interest by querying the Data Directory; they then download the selected files, along with the accompanying metadata files that provide the user with the context and assumptions under which the data were collected. Subsequent manipulation and analysis of data sets are under control of the user, using tools of their choice.

F.2 Data Sharing

Data collected with EPA funds must be made accessible (U.S. EPA 1997a; 1995b). EMAP-funded data are made available directly through the EMAP Public Web Site or through an arrangement whereby the data collectors place the data on their own web site (U.S. EPA 1997b). A hold-back period can be used before public release to allow principal investigators time to publish results.

F.3 EMAP Public Web Site

The EMAP Web pages are on the EPA public access World Wide Web server (*http://www.epa.gov/emap*). Guidelines for making data available on this site are given in Strebel and Frithsen (1995a), as updated by U.S. EPA (1997c). Data sources have a wide degree of freedom in software used (ASCII files for tabular data and Arc/Info export files for GIS data are common) and in how they structure the data sets, although certain minimal formats are encouraged (such as use of the federal Integrated Taxonomic Information System species codes). Division Director (or equivalent) approval is required to publish data on the EPA public access Web server.

F.4 EMAP Data Directory

The EMAP Data Directory lists data sets identified to be of interest to EMAP researchers. Not all of these are in the possession of EMAP. The Directory, originally based on NASA Directory Interchange Format, or DIF (NASA 1991), has been updated to comply with FGDC standards (FGDC 1994) so that the Data Directory and Data Catalog together meet the FGDC contents standard. EMAP guidelines are given in Frithsen and Strebel (1995) and Frithsen (1996a, 1996b), as updated by U.S. EPA (1996b).

F.5 EMAP Data Catalog

The EMAP Data Catalog provides information about the data, such as a description of the methods used or notes on data quality. Guidelines for preparing catalog metadata are given in Strebel and Frithsen (1995b) and Frithsen (1996a), as updated by U.S. EPA (1996c). EMAP does not place any data set on the Public Web Site without the accompanying metadata.

F.6 Further Information

Connect to the EMAP Public Web Site (*http://www.epa.gov/emap/*) or contact one of the following:

Stephen Hale
U. S. Environmental Protection Agency
27 Tarzwell Drive
Narragansett, RI 02882
Email: hale.stephen@epamail. epa.gov
Telephone: 401-782-3048
FAX: 401-782-3030

Melissa Hughes
OAO Corporation
27 Tarzwell Drive
Narragansett, RI 02882
Email: hughes.melissa@epamail.epa.gov
Telephone: 401-782-3184
FAX: 401-782-3030

References

Barton, G. 1996. NOAA Environmental Services Data Directory. Earth System Monitor, December 1996. 6–8.

Barton, G. 1997. NOAA and the Federal Geographic Data Committee. Earth System Monitor 7(3), March 1997.

CENR. 1994. The U.S. Global Change Data and Information System Implementation Plan. A report by the Committee on Environment and Natural Resources, National Science and Technology Council, Washington, D.C.

Chinn, H. and Bledsoe, C. 1997. Internet access to ecological information—the US LTER All-Site Bibliography Project. BioScience 47(1):50–57.

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 10, 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.

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.

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.

LTER. 1995. Draft proceedings of the 1995 Long-Term Ecological Research Data Management Workshop, July 27–29, 1995, Snowbird, Colorado.

NASA. 1991. Directory Interchange Format Manual; Version 4.0. NASA, National Space Science Data Center, Greenbelt, MD. December 1991.

NRC. 1995. Finding the forest in the trees: The challenge of combining diverse environmental data. National Academy Press, Washington, DC. 129 pp.

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 (EMAP), Washington, DC. Prepared by Versar, Inc., Columbia, MD.

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.

U.S. EPA. 1995. 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. 1996a. EMAP information management plan. Draft, Oct 30, 1996. U. S. EPA, NHEERL, Narragansett, RI.

U. S. EPA. 1996b. Addendum to: Guidelines for the information management directory. U. S. EPA, NHEERL, Atlantic Ecology Division, Narragansett, RI.

U. S. EPA. 1996c. Addendum to: Guidelines for the information management catalog. U. S. EPA, NHEERL, Atlantic Ecology Division, Narragansett, RI.

U.S. EPA. 1997a. 1997 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. 1997b (in prep). EMAP Research Plan (March 1997 draft). U.S. Environmental Protection Agency, ORD, NHEERL, Research Triangle Park, NC.

U. S. EPA. 1997c. Update to: Guidelines for distributing EMAP data and information via the Internet. U. S. EPA, NHEERL, Atlantic Ecology Division, Narragansett, RI.

Williams, N. 1997. How to get databases talking the same language. Science 275:301-302.

Appendix G
EPA IRM Vision Elements

This appendix outlines how the EMAP–IM system addresses the EPA IRM Vision Elements (IRM 1998).

1. **Public Access.** The EMAP–IM system components (including the Data Directory, web sites, data standards) disseminates and provides electronic access though a widely available medium (Internet). This access allows users to educate and empower themselves because the information is organized into categories (e.g., EMAP program names, geographic regions, sample media, information product type) that deliver the full range of EMAP products, including data, metadata, and publications. EMAP–IM information will become even more accessible as system components are made more compatible with tools and standards of other relevant groups such as CENR through application of Web technology, and Z39.50 and FGDC standards.

2. **EPA Access.** The EMAP–IM system components are available to all EPA employees through existing desktop technology. They supply EMAP researchers and EPA managers with technical tools, guidance, resources, and interfaces to access EMAP information for decision making and research planning. All categories of EMAP information are available through this access mechanism to provide an information library based on commonly understood features of the environmental monitoring program.

3. **Data Integration.** Indexing of data sources and metadata in the EMAP Data Directory supports comprehensive environmental protection by increasing user access to available data. EMAP's adoption of standard directory tools (e.g., Z39.50, Global Change Master Directory keyword identifiers) recommended by CENR will support this access. The use of standard World Wide Web technology that is currently implemented by many other leading environmental information programs (e.g., LTER, FGDC) will ensure adequate reporting, organization, and display of available EMAP information to potential users.

4. **Environmental Information.** The availability of EMAP information through the EMAP Data Directory and EMAP Public Web Site support the EMAP mission, i.e., to collect data to fill gaps and to share useful data collected by others by tracking externally managed information and allowing exchange of data among research partners. This availability will

allow program managers to determine requirements for data collection (data gaps) and identify existing useful information resources.

5. **Solid IRM Foundation**. The EMAP–IM system is based on a solid IRM foundation endorsed by EPA (i.e., Web-based indexing of distributed data) that is flexible and robust enough to efficiently meet the evolving mission and needs of EMAP. Open participation in planning this system is ensured by review of the Information Management Plan by the Review Team, the IMWG, SIRMO/OIRM, and the Working Groups. Recommended data standards for EMAP data sources that are in accordance with EMAP and CENR standards provide a foundation for delivery of quality data and metadata. Maintenance of the Data Directory by the experienced EMAP–IM (AED) staff will ensure the integrity of information tracking (e.g., maintaining accurate knowledge of the quality, location, and access methods for distributed data).

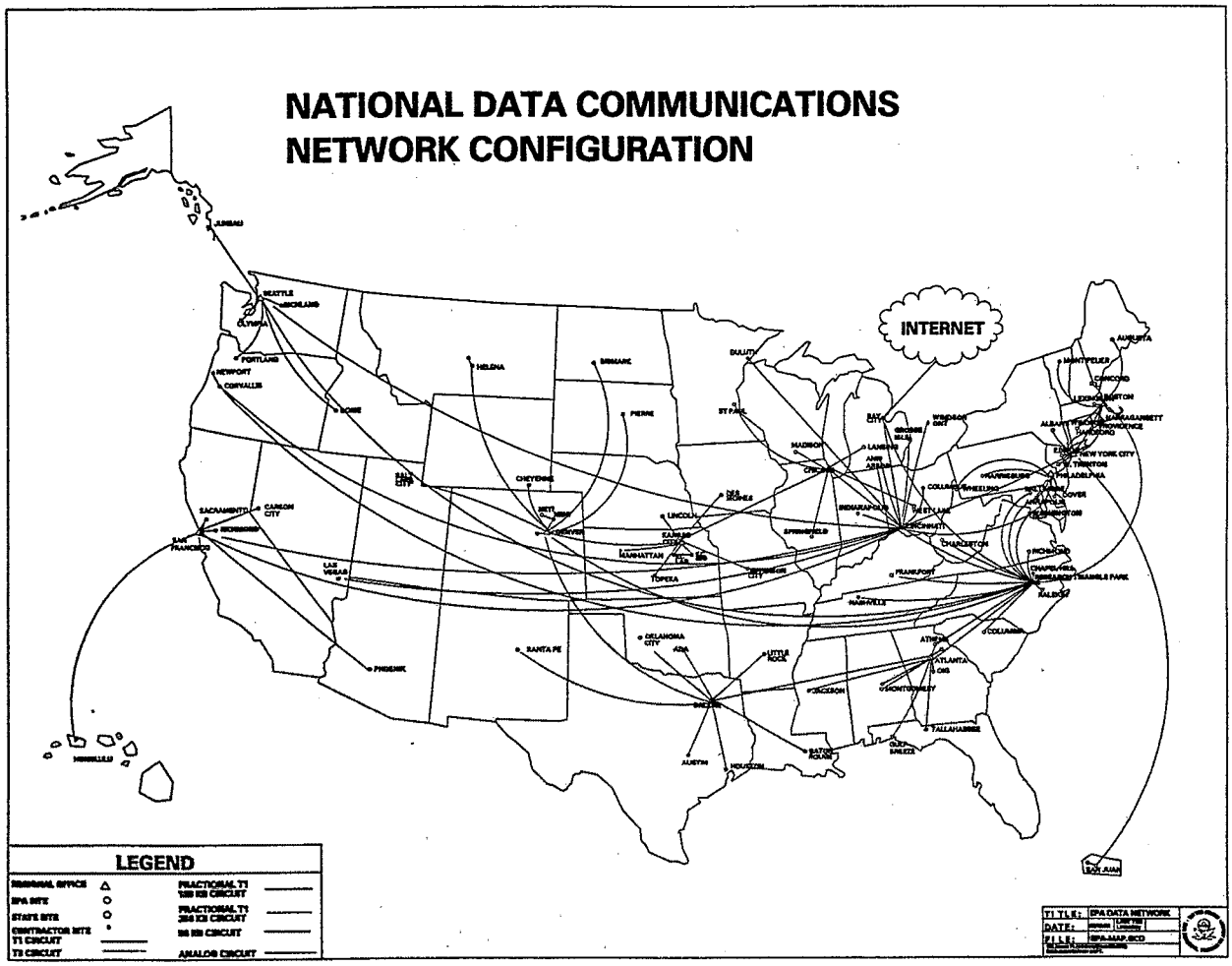
6. **Reduce Reporting Burden**. The Web-based EMAP–IM system reduces reporting burden by allowing EMAP to reference information that is already maintained at other locations. The EMAP Public Web Site reduces reporting burden by making EMAP publications available in electronic formats that users can read and download without making requests to EPA offices. The EMAP–IM system offers opportunities to automate many of the processes associated with data delivery; currently, data collection with electronic instrumentation is automated, and EMAP is adding electronic tools for creating metadata and Data Directory entries, and for reporting of data and metadata updates.

7. **Electronic Management**. The EMAP–IM system is adopting the latest Web-based tools to allow its users to conduct all transactions electronically. The EMAP Public Web Site allows users to access data without requiring delivery of printed reports, mailing of materials.

8. **Communications**. The system primarily allows users to find the information they need without intervention by data managers. Transmitting valuable information about completed data sets from the original researchers to the users, and cross-referencing information in related disciplines and geographic regions to improve cross-discipline uses of data will greatly improve the system. In addition, the EMAP–IM system's capabilities for "people-to-people" electronic communications is being enhanced. The ability of researchers to exchange data electronically is now limited by the lack of a site accessible to non-EPA research partners; this deficiency results in time delays and additional data conversion efforts that hinder research efficiency. Plans are in place to improve data exchange mechanisms for these non-EPA partners.

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

Figures on the following pages show the configuration of the EPA Wide Area Network that is referred to in Section 4.5 (System Configuration), and the hardware at the Atlantic Ecology Division where the EMAP–IM system is housed.



Draft: EPA/EMAP Computing Environment

Figure H-1. Intelligent Information Network Deployment

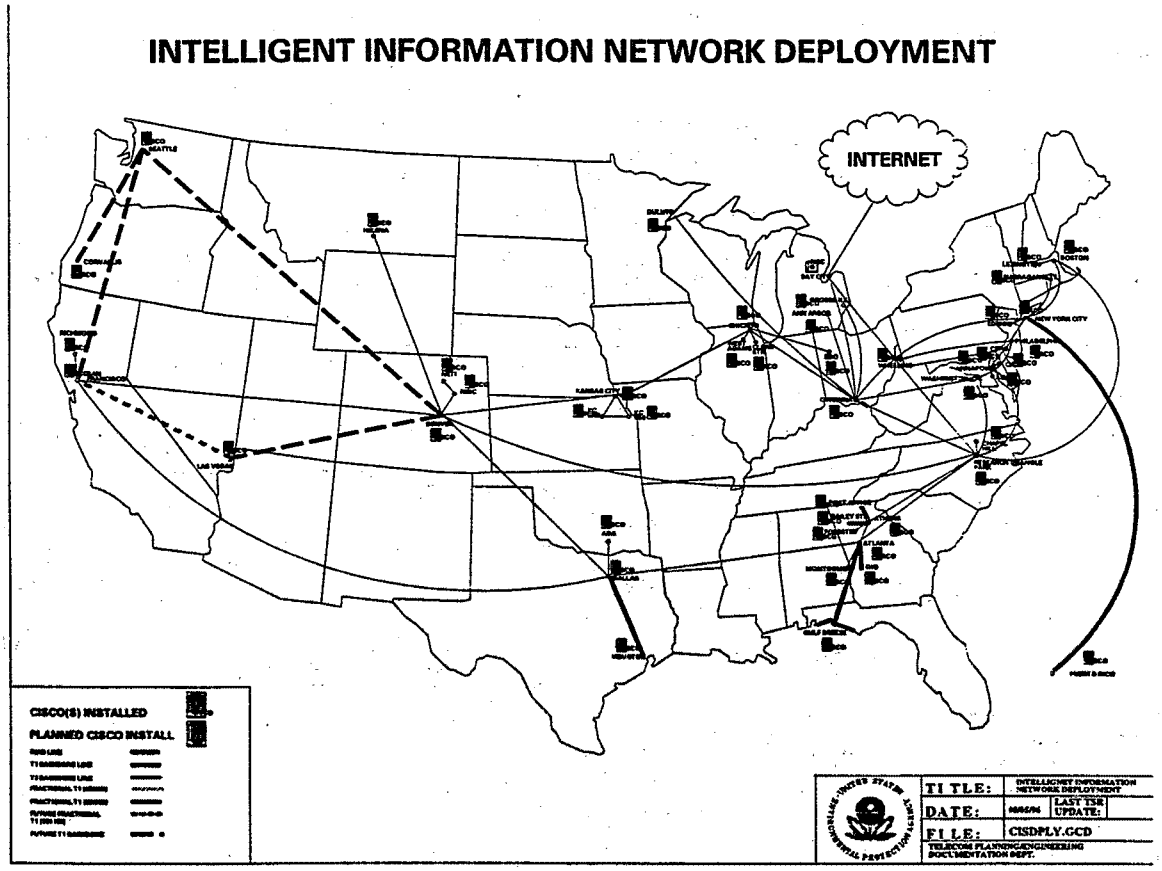


Figure H-2. National Data Communications Network Configuration

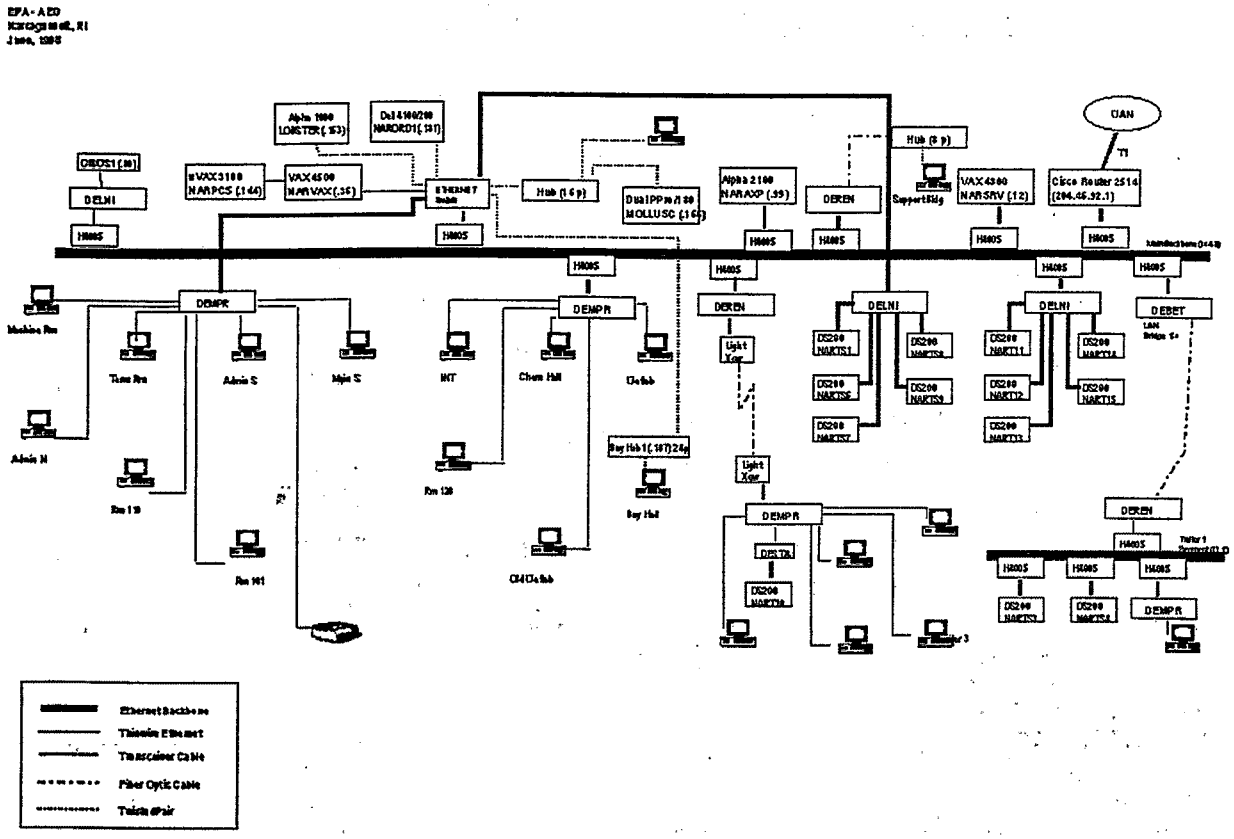


Figure H-3. EPA (AED) Hardware Architectural Design

Appendix I
EMAP Archival Plan

I.1 Introduction
I.2 Requirements for EMAP Data Storage and Usability
I.3 Types of Data Comprising EMAP
I.4 Current Digital Data Backup/Archival Scheme
I.5 Long-Term Goals for Digital Data Archives
I.6 EMAP Digital Archive Tape Validity Testing
I.7 Migrating EMAP Data to New Hardware and Software
I.8 EMAP Archival Tracking System

I.1 Introduction

This section of the EMAP Information Management Plan addresses the long-term archival of media for the Environmental Monitoring and Assessment Program (EMAP). Media are defined to include any information that is considered to have inherent value to EMAP, which may be categorized into both analog and digital data. EMAP data are stored at the Atlantic Ecology Division (AED) data center in Narragansett, Rhode Island, and include a variety of physical documentation, inventories, data, metadata, and other material. A subset of these data is stored at EPA's National Computer Center at Research Triangle Park in North Carolina (RTP). EMAP digital data includes all data and processes needed to support storage, search, and data retrieval.

Described in this plan are the data storage and usability requirements for EMAP, the types of EMAP media, and AED's archival and tracking plan. Under development is the EMAP Archival Preservation and Tracking System (EAPTS), a new system that is functionally superior to the current system. This system shall address the preservation of analog and digital archives, the media validity testing, and the tracking of the media life cycle from archive creation through validity testing.

I.2 Requirements for EMAP Data Storage and Usability

EMAP monitoring data are important enough to have a separate backup and archival plan, exclusive from data supporting other endeavors. EMAP data, metadata, documents, and applications must be usable decades from now, even as changes occur in hardware, software, and operating systems. Data must be made available when needed, and must be made fully recoverable in the event of online storage failures. In addition, all data that supports EMAP, including documentation and supporting software applications, will be maintained and tracked.

I.3 Types of Data Comprising EMAP

Much of the existing analog data is in the form of design documentation originating from EMAP Central (Phase 1) Information Management. This data will be integrated into the current AED EMAP archives. As shown in Table I-1, digital data supporting EMAP include image; varieties of textual information; proprietary spatial data formats such as ARC/INFO coverages; SAS and Oracle database programs, data, and the metadata that describes the logical and conceptual structures of records; and web media containing images, ASCII data sets and HTML files.

Table I-1. EMAP Media Types on AED Storage Devices

Operating System	NT	NT	UNIX
Data Stores	SAS	ARC/INFO	Oracle
Programs	SAS, Oracle	ARC/INFO	Oracle
Web Media	(none)	(none)	Graphic, ASCII, HTML

I.4 Current Digital Data Backup/Archival Scheme

EMAP data, programs and web media are managed from the production LAN at AED. All data on the AED LAN are backed up daily to 8mm tape. Although maintained on different operating platforms, all data are centrally backed up to "save sets." These "save sets" are archived weekly and removed to an off-site location as part of AED's disaster recovery plan.

EMAP digital data are stored in data centers at both AED and RTP. Subsets of information that comprise the AED EMAP Home Page are routinely transferred as needed to the RTP public access web server. To support the plan for long-term digital data archives, a process is under development to automate the collection and transfer of the RTP EMAP exhibit (a subset of the AED EMAP exhibit) back to AED for incorporation into the EMAP monthly archive set.

No EMAP digital data shall be stored permanently on an EMAP custodian's local disk drive. All data, metadata, and applications will be stored on the LAN disks, according to EPA policy, to ensure that these data get properly backed up and archived.

I.5 Long-Term Goals for Digital Data Archives

EMAP media shall be managed independently, shall be routinely archived, and shall be routinely tested to verify that archived digital data has not degraded. In addition, in compliance with EPA policy, all official records created or collected shall be inventoried at least triennially, in order to provide a complete and comprehensive accounting of the Agency's holdings. Further, in coordination with the AED Records Management Program, all records shall be managed according to EPA's Information Resources Management Policy, as documented in the U.S. EPA Directive 2100, Chapter 10 (U.S. EPA 1996h). Accordingly, records management consists of three basic stages: creation, active maintenance and use, and disposition. The records life cycle is initiated by the creation, collection or receipt of records in the form of data or documents in the course of carrying out EPA's administrative and programmatic responsibilities. The life cycle continues through the processing and active use of the information in the record, until the record is determined to be inactive. The final step in the life cycle is disposition which frequently includes transfer to inactive storage, followed by transfer to the National Archives Records Administration (NARA) or destruction. Once the disposition of EMAP records are considered inactive, all digital records maintained under the program shall be provided to NARA for permanent retention (U.S. EPA 1996h).

Archive tape sets will be created and compared in duplicate. One archive tape set shall be stored on-site in an environment-controlled and hazard-proof area, and the other tape set shall be likewise stored at an off-site location.

The entire EMAP digital data collection will be archived monthly to a medium with an archive durability rating of 20 years or more. The media currently in use is 8mm tape. All data written to tape will be automatically verified as part of the archive process by reading and checking the data immediately after it is written. If the verify pass fails due to a tape defect or hardware error, then the problem will be resolved and the process restarted.

A tape subsystem's ability to recover data depends upon the combination of error detection and error correction techniques, including use of parity, cyclic redundancy checks (CRCs), error detection codes (EDCs), and error correction codes (ECCs). The sophistication and robustness of the methods used determine the ability to detect and/or correct. The goal at AED is to leverage state-of-the-art technology in tape error detection and error correction to maximize tape archive recovery reliability.

I.6 EMAP Digital Archive Tape Validity Testing

All practical archive systems are subject to media degradation. Consequently, the EMAP archive plan includes routine tape testing and media renewal as necessary. Each archive tape set is a candidate for routine media validity testing. Testing of tape archive sets will be done twice annually to verify readability, and to verify that there has been no degradation of the media. This procedure shall involve both tape sets in the selected archive set. Ten percent of the entire collection of archived tapes shall be tested annually. Archive set selection shall be determined randomly from two categories of tape sets; a young category and an older category of tape sets, where the older category shall be weighted heavily for selection. After the selected tape set is written back to disk, it will be verified to be exactly identical to the data on tape. If the verify pass fails, the tape will be recreated from the duplicate tape set, such that the archive set once again comprises two error-free copies.

Archive tape validity testing shall include the monitoring of tape error rate against thresholds of advertised engineering estimates. Tape error rate is defined as the average number of errors that can be expected per unit of information processed. When observed tape error rate approaches error rate thresholds, tapes will be re-written to create a new tape archive set.

I.7 Migrating EMAP Data to New Hardware and Software

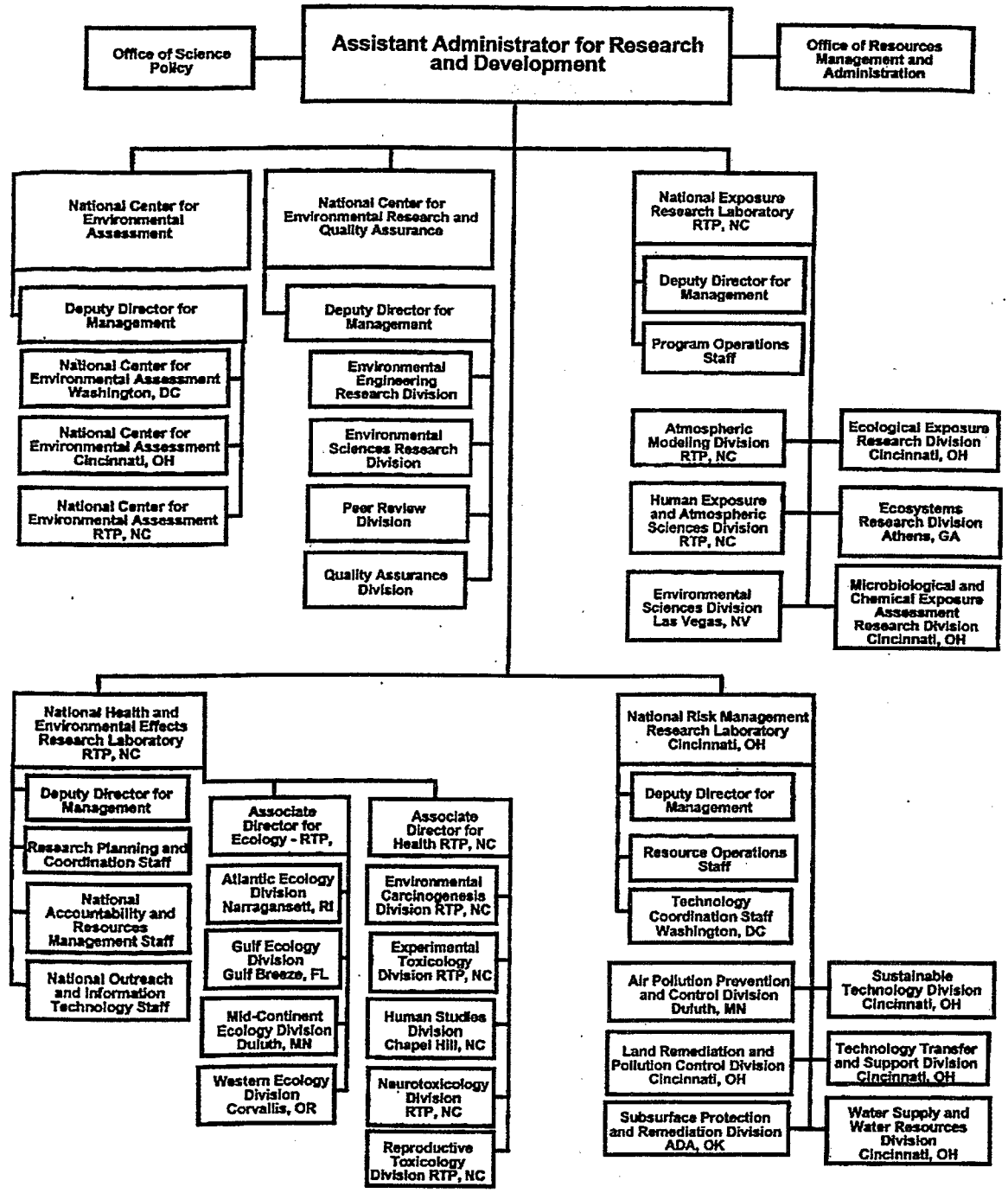
It is an inescapable fact that all EMAP digital archives will eventually need to be converted from the system originally used to create them to some new technology. This will happen because the original system will become obsolete, unsupported, and eventually unavailable in the market place. As system architectures change—hardware (e.g., servers, tape systems, etc.); software (e.g., operating systems, application programs, etc.); and backup/archive media (e.g., tapes, CDs, etc.)—all EMAP data archives shall be recreated, shall be thoroughly tested to verify readability, and shall be verified that no data has degraded. If there are architectural changes made to the system and if there are system dependencies under which EMAP and its hardware and/or software components operate, it will be recorded in the EAPTS archival tracking system database, and shall be addressed appropriately.

I.8 EMAP Archival Tracking System

A database application shall be developed to maintain, track and report on the various components of the EMAP analog and digital archives. The proposed database shall include structures sufficient to identify an analog or digital resource, including software and version dependencies, its storage location, as well as archive set hardware and software dependencies and testing history.

Appendix J
Organization of ORD Offices and Laboratories

Office of Research and Development



Appendix K
Contributors to the Development of the EMAP IM System

(in alphabetical order)

K.1 EMAP Information Management Working Group

Gary Collins	NERL, Cincinnati
Sidney Draggan	Office of the Administrator, Washington, DC
Virginia Engle	NHEERL, Gulf Ecology Division, Gulf Breeze
Pat Gant	NHEERL, Atlantic Ecology Division, Annapolis
Stephen Hale	NHEERL, Atlantic Ecology Division, Narragansett
Linda Harwell	NHEERL, Gulf Ecology Division, Gulf Breeze
Stephen Lozano	NHEERL, Mid-Continent Ecology Division, Duluth
John Macauley	NHEERL, Gulf Ecology Division, Gulf Breeze
Anne Neale	NERL, Environmental Sciences Division, Las Vegas
Tony Olsen	NHEERL, Western Ecology Division, Corvallis
John Paul	NHEERL, Atlantic Ecology Division, Narragansett
Larry Rossner	NHEERL, Atlantic Ecology Division, Narragansett
Denice Shaw	NERL, Research Triangle Park and Washington, DC
Robert Shepanek	NCEA, Washington, DC
Charissa Smith	ORMA, Washington, DC
Timothy Snoots	SC Dept. of Marine Resources, Charleston, SC
Allen Sparks	NHEERL, Gulf Ecology Division, Gulf Breeze

K.2 Contributors to EMAP Information Management, 1989–1998

We regret that this list is not complete but acknowledge the others who contributed to the EMAP Information System development:

Matt Adams	CSC
Scott Augustine	US EPA
Debra Battista	CSC, ROW Sciences
Warren Beer	US EPA
Dave Bender	OAO

Chuck Berry	CSC
Bob Booher	Dyncorp Viar
Jim Brown	Battelle Pacific Northwest Laboratories
Harry Buffum	OAO
Marlys Cappaert	OAO
Jon Clark	TPMC
Dwight Clay	US EPA
Mickey Cline	US EPA
Paul Cole	TPMC
Gary Collins	US EPA
Randy Comeleo	OAO
Lawrence Cooley	CSC
Jane Copeland	CSC, ROW Sciences, OAO
Nancy Cunningham	OAO
Sidney Draggan	US EPA
Virginia Engle	US EPA
Ted Ernst	MET
Ann Fields	George Mason University
Sue Franson	US EPA
Jeffrey Frithsen	Versar, US EPA
Donald Fulford	US EPA
Pat Gant	US EPA
Brian Goodno	CSC
Steve Greenfield	US EPA
Stephen Hale	US EPA
Linda Harwell	US EPA
Karl Hermann	US EPA
George Hess	North Carolina State University
Mason Hewitt	US EPA
Melissa Hughes	CSC, ROW Sciences, OAO
Julie Irwin	Lockheed
Robert Isaak	Battelle Pacific Northwest Laboratories
Dave James	Lockheed
Linda Kirkland	US EPA
Bruce Kissinger	Battelle Pacific Northwest Laboratories
Chuck Liff	University of Nevada at Las Vegas, US Forest Service
Jane Lovelace	TPMC
Stephen Lozano	US EPA
John Macauley	US EPA
Mark Madsen	US EPA
Doug Mann	Battelle Pacific Northwest Laboratories
Patricia Martel	OAO

Nina Mata	Bionetics, TPMC
Scott McAskill	OAO
Mary Messinger	TPMC
Kathy Moore	Battelle Pacific Northwest Laboratories
Gene Myers	US EPA
Anne Neale	US EPA
Jim Nee	TPMC
Tony Olsen	US EPA
John Paul	US EPA
Janis Peterson	OAO
Ann Pilli	CSC
Steven Rego	US EPA
Tom Richter	OAO
Victoria Rogers	Lockheed
Jeffrey Rosen	CSC, AMS and TPMC
Larry Rossner	US EPA
John Schweiss	US EPA
Tony Selle	US EPA
Walter Shackleford	US EPA
Denice Shaw	US EPA
Robert Shepanek	US EPA
Rod Slagle	Lockheed
Charissa Smith	US EPA
Timothy Snoots	SC Dept. of Marine Resources
Allen Sparks	US EPA
Donald Strebel	Versar
Jim Thomas	Battelle Pacific Northwest Laboratories
Carol Thompson	DRI
Helene F. Thoreson	Versar
Mark Tooley	North Carolina State University
Timothy Wade	US EPA
Donald Worley	US EPA

Appendix L
Partial Bibliography for EMAP IM Program

Boyd, J.L. and R.L. Slagle. 1993. EMAP Prototype core DCD system users guide. Prepared by the Lockheed Environmental Systems and Technologies Company, Las Vegas, Nevada, January 3, 1993. R-files. 5/93

Buffum, H. and S. Hale. 1997. 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. 1997. MAIA-Estuaries 1997 summary database data format and transfer guidelines. Atlantic Ecology Division, NHEERL, U. S. Environmental Protection Agency. Narragansett, RI. 33 p.

Environmental Monitoring and Assessment Program. 1991. A desktop information and mapping system for near coastal areas of the mid-Atlantic United States: Supporting EPA's Environmental Monitoring and Assessment Program. U.S. Department of Commerce, National Oceanic and Atmospheric Administration. Rockville, MD.

Franson, S.E. 1990. Data confidentiality in the Environmental Monitoring and Assessment Program (EMAP): Issues and recommendations. Revision 0, June 13, 1990. R-1854. 12/92

Franson, S.E. 1991. Proposed policy and rationale: Use of data collected under the auspices of the Environmental Monitoring and Assessment Program (EMAP). Draft July 1991. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. R-1854. 12/92

Frithsen, J.B. 1996. "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. 1996. "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.

Frithsen, J.B. and D.E. Strebel. 1995. Summary Documentation for EMAP Data: Guidelines for the Information Management Directory. April 30, 1995. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and research Program (EMAP), Washington, DC. Report prepared by Versar, Inc., Columbia, MD.

Frithsen, J.B., D.E. Strebel and H.F. Thoreson. 1992. Detailed Documentation of Data Sets for Scientific Assessments: Initial Guidance and an Example. September 30, 1992. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and Research Program, Las Vegas, NV. Report prepared by Versar, Inc., Columbia, MD.

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.

Kissinger, B. 1993. Information Management Proof of Concept, Architecture Standards. Draft, Version 1.0, February 18, 1993. Pacific Northwest Laboratories, Arlington, VA.

Mann, D.D. and R. Shepanek. 1995. EMAP Information Management Virtual Repository (Final Draft), April 1995. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program.

Mata, N. 1991. Standard operating procedures for EMAP information management technical workgroups. TS-PIC-91383. U.S. Environmental Protection Agency, Warrenton, VA.

Mata, N.J. 1993. The Environmental Monitoring and Assessment Program Geographic reference Database Development Plan. TS-PIC-93401. Environmental Monitoring Systems Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, NV 89193-3478. R-1853. EMAP-IM. 1/93.

Puterski, R., J.A. Carter, M.J. Hewitt, H.F. Stone, L.T. Fisher, and E.T. Slonecker. 1990. GIS Technical Memorandum 3: Global positioning systems technology and its application in environmental programs. Report completed for the U.S. Environmental Protection Agency under EPA Contract 68-CO-0050. U.S. Environmental Protection Agency, Las Vegas, NV. R-EMAP-GIS-BOOK, 9/92.

Rosen, J.S., J. Beaulieu, M. Hughes, H. Buffum, J. Copeland, R. Valente, J. Paul, F. Holland, S. Schimmel, C. Strobel, K. Summers, K.J. Scott, and J. Parker. 1990. Environmental

Monitoring and Assessment Program data base management system for Near Coastal Demonstration Project. EPA/600.X-90/207. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI.

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

Shepanek, R. and D.D. Mann. 1995. EMAP Information Management Standards Manual (Final Draft) April 1995. Washington, DC. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program.

Southern California Coastal Water Research Project and 11 others. 1995. Information Management Plan for the Southern California Bight Pilot Project. Southern California Coastal Water Research Project. Westminster, CA.

Strebel, D.E. and J. B. Frithsen. 1991. Handling Supporting Information for EMAP External Data Sets. December 31, 1991. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and Research Program, Las Vegas, NV. Report prepared by Versar, Inc., Columbia, MD.

Strebel, D.E. and J.B. Frithsen. 1995. Scientific Documentation for EMAP Data: Guidelines for the Information Management Catalog. Draft April 30, 1995. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and research Program (EMAP), Washington, DC. Report prepared by Versar, Inc., Columbia, MD.

Strebel, D.E. and J.B. Frithsen. 1995. Guidelines for Distributing EMAP Data and Information via the Internet. April 30, 1995. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and research Program (EMAP), Washington, DC. Report prepared by Versar, Inc., Columbia, MD.

Thoreson, H.F., D.E. Strebel and J.B. Frithsen. 1992. User requirements for Directory Interchange Formats for EMAP. September 30, 1992. Report prepared for the U.S. Environmental Protection Agency, Environmental Monitoring and Research Program, Las Vegas, NV. Report prepared by Versar, Inc., Columbia, MD.

TPMC. 1994. User Interaction and Planning Support for EMAP–IM, Technology Transfer Design Workshop Notes. February 7, 1994. ITAS8-286. Prepared for U.S. Environmental Protection Agency. Prepared by Technology Planning and Management Corporation, Durham, NC.

U.S. EPA. 1988. Environmental Monitoring and Assessment Program: The Data Set Index (DSI). Preliminary Draft, November 10, 1988. U.S. Environmental Protection Agency, Atmospheric Research and Environmental Assessment Laboratory, Research Triangle Park, NC.

U.S. EPA. 1989. Conceptual Design for the EMAP Information System. Preliminary Draft, April 11, 1989. U.S. Environmental Protection Agency, Atmospheric Research and Environmental Assessment Laboratory, Research Triangle Park, NC. R-2054. 4/92

U.S. EPA. 1990. Environmental Monitoring and Assessment Program: Information Management Committee Charter. EPA/600/X-89/000. June 1990. Environmental Monitoring Systems Laboratory, US Environmental Protection Agency, Las Vegas, NV. EMAP–IM, 4/92. Shelf.

U.S. EPA. 1990. EEI-1, Mission Needs Statement EMAP. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. R-2058. 9/98

U.S. EPA. 1990. Good automated laboratory practices: Recommendations for ensuring data integrity in automated laboratory operations with implementation guidance. U.S. Environmental Protection Agency, OIRM. Research Triangle Park, NC.

U.S. EPA. 1990. Environmental Monitoring and Assessment Program: Information Management Committee Charter. EPA/600/X-89/000. June 1990. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, NV. R-2055. 4/92

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, DC.

U.S. EPA. 1992. Summary of the Proof of Concept Joint Application Design (JAD) Session. September 25, 1992. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

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

U.S. EPA. 1993. Summary of the Proof of Concept Joint Application Design (JAD) Session II. January 15, 1993. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. EPA. 1993. IMC Project Planning Requirements. Draft, February 25, 1993. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. R-2056. 9/98.

U.S. EPA. 1993. EMAP IM POC Standards Manual. Draft, July 1993. U.S. Environmental Protection Agency.

U.S. EPA. 1993. 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. 1994. User Guide, Prototype System for Online Access to the EMAP-Estuaries Database. Version 0.2, June 15, 1994. (Draft version for test users only.) U.S. Environmental Protection Agency, Office of Research and Development, Narragansett, RI. R-2059. 9/98.

U.S. EPA. 1994. Data Catalog and Dictionary. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. EPA, 1994. User's Guide for EMAP IM System—12/8/94. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. EPA. 1995. Users guide for the EMAP IM System: Alpha Version. Draft, January 1995. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. R-Shelf.

U.S. EPA. 1995. EMAP Information Management Policies and Procedures. Draft, May 1995. Environmental Monitoring and Assessment Program. U.S. Environmental Protection Agency, Office of Administration and Resources Management, National Data Processing Division, Research Triangle Park, NC. R-2057. 9/98

U.S. EPA. 1996. EMAP Information Management Plan, October, 1996 draft. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. EPA. 1996. 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. 1996. Addendum to: "Guidelines for the information management directory," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.

U.S. EPA. 1996. Addendum to: "Guidelines for the information management catalog," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.

U.S. EPA. 1997. Environmental Monitoring and Assessment Program (EMAP): Cumulative Bibliography, 1989-1996. Draft, March 1997. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. R-2024. 5/97.

U.S. EPA. 1998. Update to: "Guidelines for distributing EMAP data and information via the Internet," U.S. EPA NHEERL, Atlantic Ecology Division, Narragansett, RI.

☆U.S. GOVERNMENT PRINTING OFFICE: 1999 - 750-101/00056