GUIDELINES FOR DISTRIBUTING

EMAP DATA AND INFORMATION

VIA THE INTERNET

Prepared for

Dr. Robert F. Shepanek
Office of Modeling, Monitoring
Systems, and Quality Assurance
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460

Prepared by

Donald E. Strebel
Jeffrey B. Frithsen
Versar, Inc
9200 Rumsey Rd
Columbia, MD 21045

April 30, 1995

The suggested citation for this report is:

Strebel, D.E. and J.B. Frithsen. 1995. 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.

EXECUTIVE SUMMARY

The Internet is a world-wide confederation of computer networks that makes possible the exchange of electronic messages, information, and data files. The U.S. Environmental Protection Agency (USEPA) has developed a public access server to provide diverse users convenient and nearly instantaneous access to data and information pertaining to the Agency and its varied programs via the Internet. The Information Management Task Group of the Environmental Monitoring and Assessment Program (EMAP) began in June 1994 to distribute data and information about the program using the Agency's public access server. Following the outline of activities presented in the EMAP Information Management Strategic Plan (Shepanek 1994), efforts to publish EMAP data and information using the public access server will expand as the transition to the Oracle relational database is completed.

The present report details procedures for publishing EMAP data using the Agency's public access server. This will be a cooperative activity. Task groups will be responsible for the contents of their sections on the server, while the common infrastructure and significant additional work entailed by publishing data and information in this way will be provided by the EMAP Information Management staff.

The public access server can be used by the task groups to provide unrestricted access to any type of information that can be encoded and transmitted in digital form. As a general principle, any data that is ready for public release is publishable. Data should be prepared suitably and accompanied by metadata sufficient to make the data understandable and usable by unknown users over a period of years.

Once the data or information has been prepared for release by the task groups, a process analogous to formal publication of scientific results will be initiated. This process has four broad phases: Submission, Editorial Processing, Formatting, and Public Release. Task group and EMAP Information Management (IM) staff have alternating roles in completing these phases, as outlined below.

Submission

The **Task Group** submits data and its associated metadata, along with appropriate publication designations, instructions, authorization, and approvals. This will normally occur through FTP transfer to a restricted access directory on the USEPA Internet server.

Editorial Processing

The **IM Staff** logs the submission and conducts appropriate quality assurance (QA) checks and reviews. If necessary, the **Task Group** revises the submission to meet publication standards. The IM Staff will not alter the content of any submission.

Formatting

The **IM Staff** posts the submission to the USEPA limited access server, including making any format conversions or inserting links (pointers) to related data sets on the server. The **Task Group** reviews this "proof" version of the publication, within a specified time frame, for errors.

Public Release

The **IM Staff** posts the submission to the EPA Public Access Server.

Data and documentation that are not ready for public release may still be made available via the limited access internal server. In this case, the overall publication process will be similar, but several of the detailed steps will be abbreviated or eliminated. For example, formal publication authorization will not be required, QA checks will be limited, and reformatting will not be performed. While the submission will be tracked, backed up, and maintained in the standard way, the public release step will not occur.

Technical and scientific support for the data and information publication effort will be provided by the EMAP Information Management staff. This support includes the activities mentioned above, along with general maintenance of the directories, files, and interface on the server, compilation of access statistics and user comments, limited support for conversion to standard formats, and assistance with technical problems encountered by users or the task group staff.

Summary of Steps Needed for the Publication of Files via the Internet. Responsibilities of Task Group and Information Support Staff Outlined		
Task Group	**IM Support Staff**	
√		Prepare ASCII formatted or WordPerfect files of data or reports to be published(See Section 5)
√		Complete submission approval form (See Table 4-1)
√		Connect via FTP to: epaaccess.epa.gov (Contact Systems Support and Operations for user name and password)
√		Choose appropriate directory for destination of files (Directory structure outlined in Section 3)
√		Transfer files and approval form via FTP
	√	Receive automatic notification of file transfer
	√	Send receipt to submitter of files
	√	Begin inventory record for file (See Table 6-1)
	√	Complete technical review of files and approval form (See Section 4.3)
	√	Convert document files to appropriate formats (ASCII, PDF) (See Sections 4.3, 5.1)
	√	Complete scientific review of files(See Section 4.3)
	√	Post files to internal EMAP WWW and Gopher server and notify owner of posting
√		Review posted files
	√	Move files to public access server (if applicable)
	√	Update "What's New" file and complete file indexing
	√	Complete inventory record for files
√		Submit requests for file updates and deletions as needed

ACKNOWLEDGMENTS

The assistance of Tom Scheitlin and Matthew Waugh (Martin Marietta GIS and UNIX Support Group) in providing information about the USEPA Public Access Server is greatfully acknowledged. Thomas Fowler, Steve Hale, Linda Kirkland, Victoria Rogers, and Robert Shepanek provided technical review of earlier drafts of this report. This report was prepared under Contract 68-DO-0093 to Versar, Inc.

TABLE OF CONTENTS

TABLE OF CONTENTS

22\epa95\112\10267-r

TABLE OF ABBREVIATIONS

ASCII	American Standard Coding for Information Interchange
BIN	Binary File
CD-ROM	Compact Disk - Read Only Memory
EMAP	Environmental Monitoring and Assessment Program
FTP	File Transfer Protocol
HTML	Hypertext Mark-Up Language
IM	Information Management
IMS	Information Management System
IMTC	Information Management Technical Coordinator
MAIA	Mid-Atlantic Integrated Assessment
NCSA	National Center for Supercomputing Applications
OMMSQA	Office of Modeling, Monitoring Systems, and Quality Assurance
PDF	Portable Document Format
SRS	Scientific Review Staff
TC	Technical Coordinator
TD	Technical Director
TSS	Technical Support Staff
TXT	Text Formatted File
USEPA	United States Environmental Protection Agency
WAIS	Wide-Area Information Server
WP5	WordPerfect 5.X Formatted File
WWW	World Wide Web

1.0 INTRODUCTION

The Environmental Monitoring and Assessment Program (EMAP) is coordinated by the U.S. Environmental Protection Agency (USEPA) and is collecting data with which to assess the general condition of the nation's ecological resources. An information management system has been developed to capture, preserve, and provide to users data collected by the program, and other data needed to complete assessments. One of the principal functions of the EMAP Information Management System is to provide EMAP science and management staff access to EMAP data, publications, and other information about the program. The system also provides access to EMAP data and information to a diverse set of users including the general public, congressional staff, government officials, and academic scientists. One approach being used to provide data and information to public users is publication using the electronic information network known as The Internet. The present document is a guide for those within EMAP responsible for preparing and submitting data for distribution via the public access server that USEPA maintains on the Internet.

The *EMAP Information Management Strategic Plan* (Shepanek 1994) specifies a requirement for a well developed capability for users to access EMAP data and information. The plan establishes a framework for developing multiple modes for user access, including:

- direct access to the EMAP data base using a custom designed interface,
- access to selected data and information through Internet, using publicly available information discovery tools, and
- publication and distribution of data and information on compact disk-read only memory (CD-ROM) and other digital information storage media.

Each of these data distribution modes presupposes the existence of an organized collection of well-documented data and associated information. The EMAP Information Management Task Group is developing a comprehensive information management system (IMS) to ensure that EMAP monitoring data are captured, preserved, and efficiently managed. The core of this system is a distributed relational data base containing both data and descriptive information (metadata) about data in the data base, a well as data organized in files external to the data base. With the data base prototyping activities nearing completion and operational implementation beginning, the next focus for attention is organizing and manipulating the data for distribution.

Direct access to the data base for selected users who are actively involved in the program is a natural outgrowth of the data base development activities, and an appropriate custom user interface has been developed in that context. Users of the EMAP IMS utilize a suite of tools to identify, locate, and describe data of interest. These tools include the data set directory that contains summary information about data (Frithsen and Strebel 1995), and

the data catalog that contains detailed descriptive information needed by scientists to understand data collected by the program (Strebel and Frithsen 1995). The Oracle Data Browser utility, as well as other tools such as SAS, are also used to directly access data in the data base.

The present report addresses the second phase of accessibility: making EMAP data and information available through USEPA's public access server. Formal data publication on distributable media such as CD-ROM will be addressed separately; however, guidelines for the preparation of such media are available (Meeson et al. 1993). Data and metadata stored within the EMAP IMS forms the core for material published through the Internet or using CD-ROM media (Figure 1-1). This data flow is consistent with the approach outlined in the Information Management Strategic Plan (Shepanek 1994) wherein EMAP data are stored in a relational database and/or separate files, and information about data and the database are stored in a virtual repository (USEPA 1994). The repository model minimizes the duplicate storage of information but increases access to the data using multiple tools. Material taken from the repository for publication through the Internet represents snapshots that will require periodic updates. As technology continues to evolve, direct links between Internet servers and the EMAP relational data base will be developed.

1.1 DATA PUBLICATION USING INTERNET

In a formal sense, EMAP inherits its charge to make data and information publicly available from legislative mandates to the USEPA to distribute data and information collected by the Agency. While providing data to users is an essential element of EMAP, this goal and the legislative mandates are reinforced by reviews of other information management systems developed and used by the Agency (USGAO 1993), the data policies adopted by the US Global Change Research Program, and the government emphasis on enhancing the electronic component of the Nation's information infrastructure. National Research Council reviewers of EMAP have specifically encouraged the program to publish data using the Internet (NRC 1994).

The Internet is a world-wide confederation of computer networks that allows the exchange of electronic messages, information, and data files. As the ability of this network to allow individual users access to vast information resources has been recognized, the use of the network has grown dramatically. Information discovery tools such as Gopher and World Wide Web have converted an anarchy of individual anonymous FTP sites into an indexed and hyperlinked knowledge base (Hayes 1994; Schatz and Harden 1994). Academic use has shown the Internet to be an effective, if somewhat informal, publication medium. Government institutions are following this lead and commercial publishers are actively planning to add Internet offerings to their repertoire. It is abundantly evident that in the near future, the network will become a major repository and delivery system for information of all kinds. Preliminary demonstrations have shown how EMAP data and information could be accessed from the USEPA Public Access Server by using commonly available information discovery tools such as Gopher, WAIS, and Mosaic.

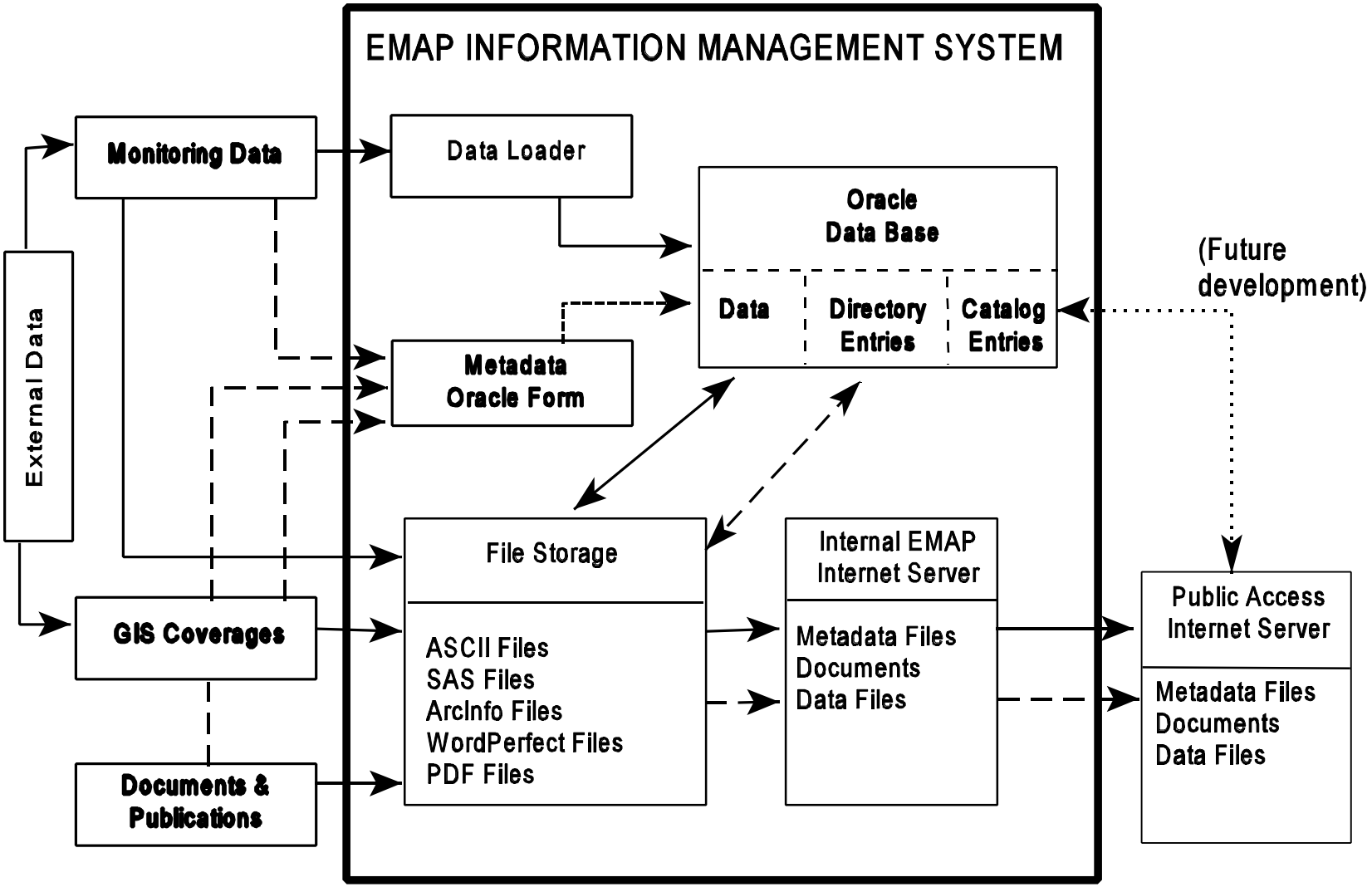
EMAP INFORMATION MANAGEMENT SYSTEM
Monitoring Data
Data Loader
Oracle
Data Base
Data
Directory
Entries
Catalog
Entries
(Future
development)
External Data
Metadata
Oracle Form
File Storage
ASCII Files
SAS Files
ArcInfo Files
WordPerfect Files
PDF Files
Internal EMAP
Internet Server
Metadata Files
Documents
Data Files
Public Access
Internet Server
Metadata Files
Documents
Data Files
GIS Coverages
Documents &
Publications

Figure 1-1. Flow of EMAP data and information into the EMAP Information Management system and out to Internet servers. Solid lines indicate data flow. Dashed lines indicate metadata flow. Dotted lines indicate future expanded pathways.

The USEPA Public Access Server allows a large number of people, via the Internet, convenient and nearly instantaneous access to data and information provided by EPA's programs. Only information that has no restrictions on use and can be distributed to the general public is posted on the server; however, as long as EMAP's total data collection is well described in the material available on the server, investigators with a need to know will be able to identify appropriate data sets and contact the USEPA for authorization to obtain them.

1.2 REPORT ORGANIZATION

The present report details organization and procedures for implementing operational public server access to EMAP data. Section 2.0 describes the types of data and information that can be provided, and discusses the related confidentiality, security, and authentication issues. A data organization scheme is proposed in Section 3.0, with particular attention to accommodating the different types of data discovery tools expected to be used. Formal procedures for publishing EMAP data and information via the server are given in Section 4.0, including adding, authenticating, revising, and removing data sets, documentation, and associated descriptive information. The formats to be used are described in Section 5.0. Section 6.0 is directed to general operational issues associated with maintaining the server as a reliable long term information resource. Implementation support to task groups (Section 7.0) and operational support to users (Section 8.0) are also described.

2.0 CONTENTS OF THE PUBLIC ACCESS SERVER

The public access server can be used by the task groups to provide unrestricted access to any type of information that can be encoded and transmitted in digital form. Ultimately, the kinds and formats of the data EMAP provides via the server will be determined by the needs of the user communities (both scientific and general public) and those elements of the EMAP data collection that it is considered of value to offer.

As a general principle, any data that is ready for public release is publishable. As a component of the EMAP Information Management System (IMS), the public success server is one means by which resource and coordination groups can meet the goal of releasing data to the IMS to support current activities including: 1) EMAP's fiscal year 1995 data analysis effort; 2) conduct of the Mid-Atlantic Integrated Assessment (MAIA) Project being carried out with Region III and other regional efforts; and 3) accessibility of EMAP-generated data and information for prospective proposers to the Office of Research and Development's new investigator-initiated grant program. Using the public access server as much as possible will reduce the cost and effort of meeting user requests for data for these and related purposes.

It is important to note that publishing the data via this mechanism is intended to be a long-term, serious commitment to provide data to a wide audience. Short-term placement of data or use of the server for immediate informal exchange between a few collaborators would be inappropriate, although the limited access server may be used in this way within the Agency. Data intended for the public access server should be prepared suitably and accompanied by metadata sufficient to make the data understandable and usable by diverse users over a period of years.

The selection of material to be posted on the public access server may require some careful consideration by the task groups. Quality assurance, review, and approval procedures are needed prior to the publication of data and documents, and the review process includes assessment of the supporting information (metadata) that will accompany data (Kirkland 1994). In addition, because the general public will have access to the server, there will always be issues of confidentiality and security to consider. Where confidentiality is an issue, it is recommended that the existence of the data not be hidden from the user, but that the usage restrictions be spelled-out and procedures for becoming an authorized user provided. Informed users with a positive attitude are more beneficial to the program than those who become convinced that there is a conspiracy to hide something.

2.1 TYPES OF INFORMATION

The expected users of data and information placed on the public access server are EMAP and other USEPA scientists; USEPA regional personnel; environmental or scientific staff from other Federal Agencies; and members of non governmental organizations concerned with

environmental issues. If fully populated with EMAP data and publications, the server can be a primary means by which these users can find information about EMAP. To be effective, the information on the server should describe: the purpose and organization of the program, environmental monitoring activities of each resource group, and regional research projects. This information should range in a logical way across various levels of detail and technical complexity to be useful to users of diverse scientific sophistication and familiarity with the program.

Examples of the kinds of information that might be on the server include:

- ASCII text files describing the program or its components.
- Published program documents and reports in word processor formats.
- Scanned images of organization charts or results graphs.
- Scanned photographs of sampling gear or monitoring sites.
- Digital files containing sound clips or video images.

Among the information files provided should be a selection of formal EMAP publications and a bibliography of significant published studies using EMAP data. Documents published by the program are part of the public record and their general availability will contribute to both public understanding and support for the program. Also particularly useful will be the annual statistical summaries, reports to Congress, and other formal output that EMAP produces to achieve its goal of assessing and reporting the condition of the Nation's ecosystems.

The technology currently exists to include digitized photographs, digital sound, and video files to supplement published reports and data. Such multi-media products may be especially useful as training tools to investigators in remote locations within the United States or internationally that have an interest in using sampling or analytical approaches developed by the program.

2.2 TYPES OF DATA

The contents of the public access server can be viewed and downloaded by anyone with access to the Internet; therefore, only data that has been cleared for public distribution will be posted on the server. Within those constraints, all levels of data might be made available, including:

- Summaries of regional conditions
- Indicator data for individual sampling stations
- Averaged station data
- Replicate data from individual sample stations
- Individual instrument measurements (e.g., CTD casts)
- Chemistry data and laboratory analyses

The EMAP Information Management System (IMS) will be the conduit for EMAP data published using the public access server (see Figure 1-1). Through the EMAP data directory (Frithsen and Strebel 1995), the EMAP IMS references both data in the relational tables in the Oracle data base management system that is the core of the EMAP IMS, and external files. Until the directory is completed by all resource groups, some data published using the public access server will not be fully referenced in the EMAP IMS.

2.3 ACCESS ISSUES

The widespread access to public information on the Internet raises issues that might not normally be considered in collegial scientific data exchange. These issues include:

- A potential audience of enormous breadth of backgrounds and degrees of scientific understanding
- Inadvertent misuse of data due to inadequate or incomplete documentation
- Failure to credit data sources appropriately due to lack of scientific training

To some extent, releasing the data into the public domain means being prepared to accept some abuses (e.g., ideological use of selected data items out of context). The best defense against unjustified criticism or misuse, however, will be careful preparation and documentation of all data, with special attention to educating the audience in the process.

While the expected users of the system (see p. 2-1) are the primary audience, consider that the following potential users may easily find and access the server:

- Academic scientists writing proposals to research environmental issues.
- Foreign scientists designing their own environmental monitoring programs.
- Journalists seeking background information on environmental disasters.
- Regional environmental study groups analyzing local monitoring data.
- College students writing term papers on environmental issues.
- Curious laymen attracted by the program name.

We do not want such users to carry away a negative impression; although perhaps they represent only isolated special cases, their impressions could have weighty impacts.

Bearing the above in mind, task groups will select which data and information in the EMAP IMS will be published using the public access server. This selection will depend on USEPA and EMAP policies, internal task group criteria, and the potential for interactions with data and information being added to the server by other elements of EMAP. In general, a task group should develop procedures for deciding about posting material to the server, rather than relying on ad hoc decisions. A consistent procedure will be easier to document, be maintainable through personnel changes, and be more efficient.

Some of the considerations for publishing data include:

- Only verified, validated data can be published
- Only data approved for release by a Technical Director (TD) or Technical Coordinator (TC) can be published
- Data to be published should respond to user needs
- Data requests and previous access history should be used to set priorities for data and documents to be selected and prepared for publication.

The above list touches on only a few of the more obvious factors bearing on whether a particular data set should be published and when. There are also a host of details that will depend partially on the scientific analyses and uses of the data. Should the data be organized by sites or years? Should replicate information be included, or only average values reported? Should all sites be reported on an equal basis, or should the focus be on the probability-based sample sites? These decisions are best made by scientists and information management staff in the individual EMAP resource groups; however, as a general guideline, resource groups and the program are best served by providing as much data as is possible through the public access server.

2.3.1 CONFIDENTIALITY ISSUES

At the present time, data distribution policies and approval for the distribution of data are the responsibility of individual resource group technical directors. Many questions that users of the public access server might raise will be avoided if EMAP establishes a uniform policy applicable to all resource groups. This policy should include a review of the types of data that will be considered confidential by the program, general guidelines for the review of data prior to public distribution, and a general rule for any time periods granted for the exclusive use of data collected by individual investigators. This would augment the information pertaining to the review and clearance of reports and data outlined in the EMAP Quality Management Plan (Kirkland 1994).

Serious thought should be given to the release of a data set in which the locational information is hidden. Unless this circumstance is clearly explained, a user who downloads the data set could conclude that the data set was incompetently collected or analyzed, and that published conclusions based on it are likely in error. If this is a possibility, it would be better not to publish the data set for general use.

This approach may be particularly appropriate in those circumstances where providing locational information might compromise the integrity of the EMAP sampling design (Franson 1991). Release of spatial coordinates may also compromise interagency agreements with cooperating partners that are coordinating other related monitoring efforts (e.g., the Forest Health Monitoring Program of the Department of Interior's Forest Service, or the Department of Agriculture's Annual Agriculture Survey). Users with a verifiable need to access this kind of data can be authorized to obtain it via more secure mechanisms.

2.3.2 Locational Data Confidentiality

Considerable debate has already been devoted to the release of information providing the exact spatial location of sites sampled by EMAP. While there are certainly valid concerns about making such information widely available, most serious users of environmental data expect such information. No spatial analyses or cross data set integration are possible without at least some degree of locational information. The substance of the debate is what degree is appropriate. For this reason, it is suggested that some locational information accompany every sampling site based data set to be distributed through the public access server.

In those cases where release of the data set is highly desirable, but where the public distribution of sampling locations is thought to compromise the sample design or conflicts with interagency agreements, spatial coordinates can be provided with specifically degraded locational precision. Thus, each sampling location should be specified by both a coordinate and an accuracy, with a footnote indicating that the original data was collected with greater accuracy. In the simplest case, for example, latitudes and longitudes could be provided to the nearest degree (e.g., 40° ± 1° N, 92° ± 1° W), which only locates a point to within about 100 km. A scientifically more satisfying algorithm would be to define an acceptable "closeness", say 10 km, and deliberately introduce a random perturbation in each coordinate of between 0 and this amount. The coordinates would then be reported with full precision, but with an (in)accuracy determined by this perturbation (e.g., 40° 35' 21.98" ± 5' N, 91° 47' 53.15" ± 5' W). The statistical properties of the set of points would then remain the same and, if the statistics of the random perturbation were also reported, precise error estimates could be given for the effects of the perturbation on any subsequent analyses.

3.0 SERVER ORGANIZATION

The EMAP data and information on the public access server will be organized as sets of files within a hierarchical directory structure. Some of the guidelines for creating this structure are:

- There will be parallel directory structures for different task groups, to maintain programmatic consistency and ease of navigation.
- Since elementary browse tools (e.g., Gopher) display one screen page at a time, there should be no more than 12 to 14 entries per directory.
- To control update issues, no information that can be linked from another location will be duplicated on the public access server.

Within each directory, the actual data will be determined by and provided by the relevant task group through the EMAP IMS. It would be useful, however, to follow an organizational scheme that includes the following:

- A text file describing the sample sites for each observation year.
- Each year's data split into separate files by geographic region.
- A text summary file providing an overview of the data for each group of files.

It is important to users that the structure of the information pertaining to individual resource groups be similar. The directory structures outlined in this chapter provides guidance for a consistent organizational structure for each EMAP Task Group, while also providing flexibility for each resource group to organize data to best meet the needs of their users.

3.1 LINKS BETWEEN ACCESS TOOLS

EMAP data and information will be distributed through the USEPA's Gopher and World Wide Web servers. To the extent possible, task groups will not be asked to make multiple submissions of the same data or information. The objective is to extract data from the EMAP IMS to the public access server once, but to do so in a way that it can be accessed by multiple Internet discovery and retrieval tools. Linking files to both Gopher and World Wide Web servers is fairly straightforward and will involve no additional steps on the part of the EMAP Task Groups.

Wide-Area Information Server (WAIS) software will be used to index text files included on the public access server. The index provides users with the capabilities of identifying

documents containing specific keywords of interest. Technical staff maintaining the public access server will complete indexing of specified files for the EMAP Task Groups.

More elaborate links between World Wide Web and WAIS servers will need additional research and are not planned at this time. It is possible to combine the two types of server and use World Wide Web Mosaic software as a front-end client with WAIS as a back-end server; however, this capability is not yet fully developed and ready for implementation.

The large growth in the number of World Wide Web servers, and the number of users with access to the National Center for Supercomputing Applications (NCSA) MOSAIC client or Netscape's client software, suggests that in the very near future, the majority of users accessing the public access server will utilize World Wide Web clients. Despite the success of World Wide Web, there is a commitment on the part of the Agency and EMAP IM to continue to design for and support Gopher clients. MOSAIC and Netscape are fundamentally graphical tools for network information retrieval, display, and query. Use of these clients typically requires good communication bandwidth. Gopher provides access to screens of text and can be used where communication links are not optimal. Additionally, the hierarchical organization of files is typically more apparent using Gopher compared to World Wide Web clients. Understanding the relationship between files is often helpful to users who have little familiarity with a program.

3.2 DIRECTORY ORGANIZATION

An organization for directories on the public access server that will contain EMAP data and information has been developed and outlined below (Table 3-1). The directory organization pertains to the Gopher server; however, World Wide Web client software will link to the files in these directories, or to additional files that have been converted to hypertext mark-up language (HTML).

Table 3-1. Gopher Directory Structure

EMAP Directory and Primary Subdirectories

Environmental Monitoring and Assessment Program
- * Welcome and overview (TXT)
- * What's New (TXT)
- * Program Overview...
  - ** Abstract and Table of Contents (TXT)
  - ** ASCII Text for Chapters...
    - *** Introduction (TXT)
    - *** History (TXT)
    - *** Ecological Risk Assessment (TXT)
    - *** EMAP: Monitoring for Results (TXT)
    - *** EMAP's Integrated Approach (TXT)
    - *** Reporting Results (TXT)
    - *** Measures of Success (TXT)
    - *** Interagency Cooperation and Partnerships (TXT)
    - *** Remaining Challenges (TXT)
    - *** Glossary (TXT)
    - *** References (TXT)
  - ** Portable Document Format (PDF) Chapters...
    - *** Introduction (PDF)
    - *** History (PDF)
    - *** Ecological Risk Assessment (PDF)
    - *** EMAP: Monitoring for Results (PDF)
    - *** EMAP's Integrated Approach (PDF)
    - *** Reporting Results (PDF)
    - *** Measures of Success (PDF)
    - *** Interagency Cooperation and Partnerships (PDF)
    - *** Remaining Challenges (PDF)
    - *** Glossary (PDF)
    - *** References (PDF)
- * Data Overview
  - ** About EMAP File (TXT)
  - ** List of EMAP Data Sets by Name (TXT)
  - ** Keywords for EMAP Data Sets (TXT)
  - ** Searchable Index of EMAP Data Set Directory (INDEX)
- * Publications...
  - ** About EMAP Publication List (TXT)
  - ** EMAP Publication List (TXT)

Table 3-1. Continued

* Personnel and Contacts...
  ** About EMAP Personnel and contacts (TXT)
  ** EMAP Personnel and Contacts (TXT)
* Resource Groups...
  ** About EMAP Resource Groups (TXT)
  ** Agricultural Lands...
  ** Estuaries...
  ** Forests...
  ** Great Lakes...
  ** Landscape Ecology...
  ** Rangelands...
  ** Surface Waters...
* Coordinating Groups and Activities...
  ** About EMAP Coordinating Groups and Activities (TXT)
  ** Air and Climate...
  ** Assessment and Reporting...
  ** Design and Statistics...
  ** Indicators...
  ** Information Management...
  ** Landscape Characterization...
  ** Methods...
* Regional Initiatives and Programs...
  ** About Regional Initiatives and Programs (TXT)
  ** Mid-Atlantic Integrated Assessment Project...
  ** Great Lakes Regional Initiative...
  ** Pacific Northwest Regional Initiative...
  ** South Florida Regional Initiative...
  ** USEPA Region I Projects...
  ** USEPA Region II Projects...
  ** USEPA Region III Projects...
  ** USEPA Region IV Projects...
  ** USEPA Region V Projects...
  ** USEPA Region VI Projects...
  ** USEPA Region VII Projects...
  ** USEPA Region VIII Projects...
  ** USEPA Region IX Projects...
  ** USEPA Region X Projects...
* Quality Management...
  ** About Quality Management (TXT)
* Cooperating Groups...
  ** About EMAP Cooperating Groups (TXT)
  ** List of EMAP Cooperating Groups (TXT)

Table 3-1. Continued

* Document and Data Formats and Software for EMAP Directories...
  ** Formats for EMAP Documents (TXT)
  ** About Portable Document Formats and Software Viewers (TXT)
  ** Formats for EMAP Data (TXT)

Subdirectories for Resource Groups (Example shown for Agricultural Lands)

*** About Agricultural Lands Resource Group (TXT)
*** Resource Group Contacts (TXT)
*** Resource Group Publication List (TXT)
*** Resource Group Publications...
  **** About Resource Group Publications (TXT)
  **** Searchable Index for Resource Group Publications (INDEX)
  **** Publication #1...
    ***** Abstract and Table of Contents (TXT)
    ***** ASCII Text for Publication Chapters...
      ****** Chapter 1 (TXT)
      ****** Chapter 2 (TXT)
    ***** WordPerfect 5.1 Formatted Chapters...
      ****** Chapter 1 (WP5)
      ****** Chapter 2 (WP5)
    ***** Portable Document Format (PDF) Chapters...
      ****** Chapter 1 (PDF)
      ****** Chapter 2 (PDF)
*** Resource Group Data...
  **** Organization of Resource Group Data and Metadata(TXT)
  **** Searchable Index of Resource Group Metadata Files (INDEX)
  **** Data Set #1
    ***** Metadata File (TXT)
    ***** ASCII Formatted Data File (TXT)
    ***** SAS Export Format Data Set (BIN)

Subdirectories for Coordinating Groups (Example shown for Information Management)

*** About Information Management Coordinating Group (TXT)
*** Coordinating Group Contacts (TXT)
*** Coordinating Group Publication List (TXT)
*** Coordinating Group Publications...
  (subdirectories should mimic those for Resource Group Publications)

Table 3-1. Continued

Subdirectories for Regional Initiatives and Programs (Example shown for MAIA)

*** About Mid-Atlantic Integrated Assessment (MAIA) Project (TXT)
*** MAIA Contacts (TXT)
*** MAIA Publication List (TXT)
*** MAIA Publications...
(subdirectories should mimic those for Resource Group publications)

4.0 PROCEDURES

4.1 INTRODUCTION

The public access server will be a dynamic system with a lifetime of several years. As EMAP data and analyses mature, the data suitable for immediate public access will change; a continual cycle of data additions, revisions, and removals is anticipated. There must, therefore, be a consistent set of procedures, followed through time, that maintain the integrity, currency, and authentication of the online data. Without this, EMAP and the Agency will lose credibility.

The procedures presented below outline the specific steps to be followed when adding, revising, or deleting files in the EMAP directories on the public access server. The primary principle is that individual task groups are responsible for their own data. The procedures, however, are designed to be consistent with Agency policy and reflect operational procedures used by the technical support groups at the USEPA's National Computer Center. The procedures below are intended to be used by EMAP Information Management Staff, Task Group Information Management Staff, and others directly concerned with the content of the server. Procedures concerning user access, routine operational maintenance, and technical support are discussed elsewhere.

All data and information that are placed on the public access server will first be staged to the internal (limited access) server. Thus, these procedures can also be used to populate the internal server. However, for data and information that are not intended for public release, the process can be streamlined by skipping several of the steps.

4.2 AGENCY POLICIES AND PROCEDURES

The Director of OMMSQA has established a goal for the EMAP Resource and Coordination groups to make their data and information (that have been verified, validated, and subjected to quality assurance requirements) available for population of the EMAP Information Management Systems (IMS) by early 1995 (Bills 1994). There are a number of programs and projects that require ready availability of these data. One mechanism for meeting this goal in a timely manner is posting the data on the public access server.

Agency procedures for publishing data and information on the public access server are being established. The current draft version of these procedures (Waugh and Weaver 1994) has been reviewed in preparing the suggested EMAP-specific procedures below. The draft agency submission procedures also draw upon the draft version of the document "Information Security and Internet Services", EPA Report 691/001, which requires appropriate certification of data being placed on the public access server.

The EMAP Information Management Technical Coordinator (IMTC) will provide appropriate certification to meet agency requirements. The functions required to establish this certification, including evaluating and addressing security and confidentiality requirements, have been delegated by the IMTC to the technical directors and technical coordinators for data and information for which they are responsible. The procedures given below establish sufficient mechanisms to meet these certification requirements.

In addition to certifications, Agency policies and procedures require that appropriate documentation be provided for each data set distributed using the public access server. This is consistent with the guidelines in the EMAP Quality Management Plan (Kirkland 1994) that specify that no data are distributed without appropriate documentation (metadata). Guidelines for the preparation of metadata are provided in Section 5.0 of this report.

4.3 PUBLISHING DATA AND INFORMATION - PROCEDURES

Placing data and information on EPA's public access server is a form of publication. As in any publication process, there must be procedures for submitting the material, reviewing it for consistency with format and content standards, making sure that it poses no security questions and is approved for release, and abstracting and advertising it so that users can find it. There will be three teams of people involved in completing this process. The first will be the task group staff who prepare the data and documentation and submit it. Once submitted, the data will be published through coordinated efforts of the EMAP Information Management Technical Support Staff (TSS) and Scientific Review Staff (SRS).

There are four generic phases to publication: Submission, Editorial Processing, Formatting, and Public Release. Each of these phases has steps to be completed in data publication that are analogous to steps in the publication of other material. The Editorial Processing phase includes such QA reviews as are necessary to establish uniform standards of content, presentation, and credibility for EMAP data and information on the public access server.

The four phases comprise the following specific sequence of steps (Figure 4-1):

Submission

1) Submit the data and information

   Directories to receive files will be created on the internal Internet server (epaacess.epa.gov), with a directory structure identical to that for the Public Access server. Task groups will normally select an appropriate directory and FTP their new data submission to it. Ideally, data files will be created from data within the EMAP IM distributed data base, or represent files documented using the EMAP data directory; however, until the data base and directory are fully populated, data files may be submitted from the task groups directly.

PROCEDURE FOR PUBLISHING EMAP DATA AND INFORMATION ON THE EPA PUBLIC ACCESS SERVER

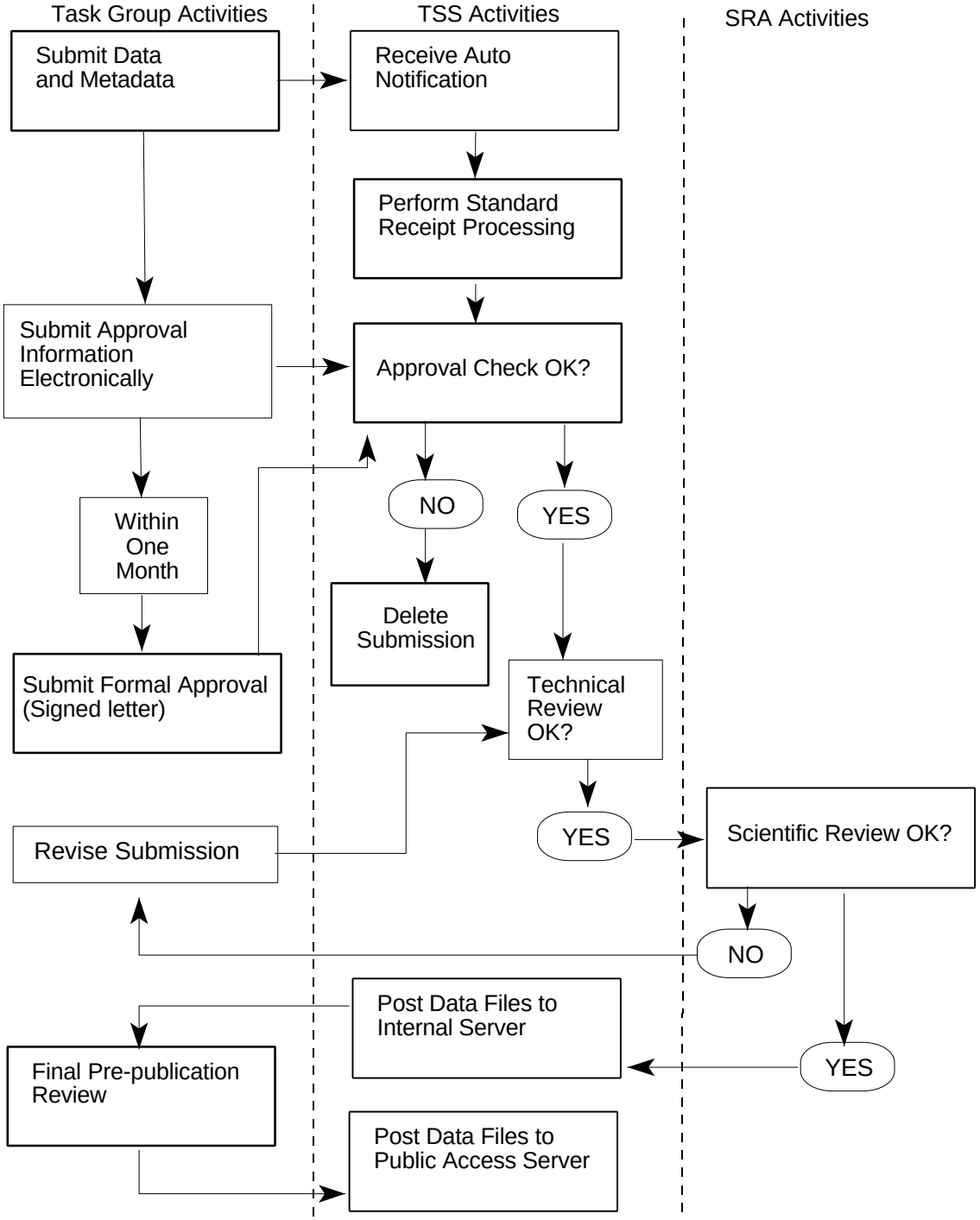

Figure 4-1. Steps for publishing EMAP data and information on the USEPA public access server

Editorial Processing

2) Notification of submission

An automatic procedure will regularly examine the directories for new submissions and send a notification message to the EMAP IM Technical Support Staff.

3) Receipt

The assigned technical support staff will backup the files submitted, inventory them in a submission tracking system, and send a receipt notification message to the submitting task group.

4) Approval Check

All data placed on the public access server must be approved for dissemination by an appropriate USEPA/EMAP official. Consistent with Agency policy, the EMAP IMTC is responsible for overall approval of all files containing EMAP data or information placed on the public access server. The EMAP IMTC has chosen to delegate that responsibility to individual technical directors and coordinators. The delegation of approval authority provides for accountability by those that are most familiar with individual data and documents and is consistent with guidelines outlined in the EMAP Quality Management Plan (Kirkland 1994). Although Agency policy concerning the public access server indicates that approval for each individual file is not necessary, provisions for individual file approval are included in the procedures outlined here to facilitate building the file inventory and tracking future file changes.

To speed processing, the initial approval can be provided by an electronic note submitted with the data, and sufficiently specific to clearly identify the data set uniquely and establish the date of the approval, approving official, and any expiration date or other restrictions on dissemination of the data. The destination (internal or external server) and any specific publication instructions should also be noted. An example is provided in Table 4-1. Because of the difficulty of electronic verification, any electronic notification of approval must be followed by a hard copy. Both hard copy and electronic file will be maintained by the TSS in an appropriate archive and tracked in the submission tracking system data base maintained by the TSS.

If an appropriate approval is not received, the technical staff notifies the person submitting the data and the appropriate approval official. Further processing of the submission is delayed until the approval is received. If no approval is received within a specified period (e.g., a month), the submission is rejected and the data file(s) removed from DGGIS. The submission tracking record is then closed. The process must restart from Step 1 if the task group wishes to attempt to publish the data at a later time.

Table 4-1.

SUBMISSION APPROVAL FORM

Instructions: Edit this form as required and place a copy in the directory along with the data submitted.

Date:

Task Group:

Person Submitting Data:
Email Address:

Person Authorizing Data Submission (TD/TC) or revision:
Approval Date:
Expiration Date:
Restrictions on Disseminations:

Is data set to be put on public access server*:
or on limited access server only:

Is this a new data set:
If not, what data set does it replace:

Path and name of files submitted:

* For data and information to be released to the public access server, the authorizing person should sign a printed copy of this form below and forward it within 30 days to:

(Sign here)

The rejection of the data submission on this basis is intended to protect the task groups and USEPA from unauthorized release of sensitive or erroneous data. The possibility of a network security breach resulting in the introduction of spurious data files that would compromise the program or the functioning of the public access server also requires deletion of unauthenticated submissions within a reasonable time.

5) Technical Review

The Technical Support Staff (TSS) then evaluates whether all of the technical standards for posting to the public access server have been met. (See section 5.0 for technical format and content guidelines) If not, the TSS work with the submitting task group to revise the data submission appropriately. The evaluation criteria applied by TSS include:

- Was the file placed in the appropriate directory (Task group, data set, year,...)?
- Is there appropriate metadata describing the files (File format, descriptors of file contents,...)?
- Is there any indication of file transfer problems (file indicated on approval form not received; number of lines inconsistent with header; garbled text, etc. . . .)?
- Are the files and associated metadata dated appropriately (submission date, revision dates, expiration date if applicable)?
- Are the files in the appropriate format?
- Is there a need for converting the files to additional formats?

The submission inventory record should be updated to indicate that the technical review was completed, or if necessary to identify problems found and actions taken. The inclusion of submission/revision/expiration dates in the data and metadata are critical to version control and establishing update histories for users and analysts.

6) Scientific Review

The IM Science Review Staff (SRS) reviews the data and associated documentation to establish whether the data set has unintended problems that would preclude publication on the public access server. If so, they will work with the submitting task group to revise the data submission appropriately. This review step is not meant to question the task group's judgment, but as with any scientific peer review, to provide an independent check that identifies problems and improves the final quality of the publication. As with the technical review, the

submission inventory record should be updated to indicate when the scientific review was completed and any actions taken as a result thereof.

A scientific review is less specific than a technical review, but includes items such as:

- Does the metadata describe the scientific origin and use of the data set adequately?
- Does the stated description of the data match what is in the data file (e.g., units, ranges) and is this what is expected for this type of data?
- Is appropriate Quality Control and calibration information included or referenced?
- Is there sufficient information to use the data in integrated analyses that span task groups?
- Are there any unforseen data confidentiality issues that would preclude publication (e.g., could the release of this data set identify sites hidden in another previously released data set?)?

Formatting

7) Post Data to Limited Access Server

Most data and documentation will pass through the editorial stage quickly. Editorial processing is designed to take no more than 10 to 14 days; however, additional time may be needed to convert complex documents to the appropriate formats. Files will then be posted on the limited access server in the directories and formats expected to be on the public access server. (Some format conversions will be performed by the TSS - see Section 5.0). Interface pointers (e.g., for WWW or Gopher) will be added by the TSS and the task group will be notified that the data set is ready for pre-publication review.

8) Task Group Review

Posting to the Internal server constitutes the "Galley Proof" stage of the publications process. Task group staff will have a limited time (recommended: 1-2 weeks) to review the material as it will appear on the server before it is released. This is an emergency "stop the presses" step only - failure to respond will be taken as consent to release the data as is.

Public Release

9) Post Data to Public Access Server

When the above eight steps have been completed, all requirements for public release of the data have been meet. The submitted files will be moved to the operational directories on the public access server, and the submission inventory record updated to this effect and closed. Information on the submission will be added to search indexes and other access aids as necessary. The "What's New" message will be updated appropriately.

4.3.1 Internal Server Procedure Modifications

If the data or information is not intended for release to the public access server, Steps 8 to 9 will not occur. While it would be advisable to complete all of the remaining steps, some can be abbreviated or skipped if the task group desires. These include: submitting a signed hard copy data submission form, completing the technical and scientific reviews, and reformatting. The task group staff should state in the "instructions" field the electronic data submission form what elements of the process are desired or should be skipped.

4.4 PROCEDURES FOR REVISIONS

Data revisions are a natural part of scientific inquiry. Generally, files will not be revised by technical support or scientific review staff. When a data revision is necessary, a task group should prepare the revised data set and submit it exactly as given above, with the following additional requirements:

- The submission must include a new data submission form specifying that this is a revised data set, giving suitable reference to the original data set, any special instructions as to how replacement is to be made, and the informal approval for the change. (Formal approval will still be required, as specified above).

- The original file and its derivatives (e.g., PDF, HTML files) will be archived by the TSS to maintain a trackable change history. The data tracking system will note when the file was archived and the record identification of the file submission record for the information replacing it.

- The metadata file will be resubmitted, noting the changes made, and the reason for them, as well as the revision date.

4.5 PROCEDURES FOR DELETING INFORMATION AND FILES

As data matures and the focus of the program changes, it may no longer be appropriate to maintain data on the public access server. In some cases this may occur also because the

data are proven to be seriously flawed, although this would normally be a rare event if quality control procedures have been followed.

The TSS should be notified by the approving authority for the data set that it is to be deleted. An email will suffice for initiating action, particularly if urgent action is required. However, a formal deletion letter should be submitted for the record.

Before making a deletion request, it is advisable to review usage statistics for the data set. The TSS and the SRS will provide information on the record of accesses to the data set and the scientific import and usefulness of the data set with respect to other online data sets. Should it appear that, for the benefit of the active users and/or the program as a whole, the data set should be maintained on line, the staff may recommend the data set not be deleted.

Data sets that are deleted will be archived by the TSS to maintain a trackable change history. The data tracking system will note when the deleted file was archived, and the reason for its deletion.

5.0 FORMATS

Data and information files placed on the public access server should be formatted in a way that provides the greatest degree of access. In general, this means adopting a “lowest common denominator” approach. EMAP IM staff cannot assure that all users will have timely access to the latest versions of all software and network browsers.

As a general rule, an ASCII "readme" file accompanies all data and information placed on the server. This file describes the contents and organization of specific directories, specifies paths to specific information, and is placed where it can be easily found by users. These descriptive files can be duplicated in multiple directories as is needed to provide information to the browsing user. At least one "readme" file should be included in each directory and subdirectory on the server. In general, the technical support staff will draft "readme" files for most directories; however, task group staff should create "readme" files for their own subdirectories. "Readme" files should not be confused with metadata files containing the detailed documentation for data sets.

5.1 DOCUMENTS

Documents published using the public access server may include short (less than one page) descriptive files of the contents of directory or subdirectory as well as long (hundreds of pages) reports containing embedded graphics and tables. Different approaches need to be used to accommodate publication of these different types of documents.

Short documents containing text only should be prepared as ASCII formatted text files. The ASCII text file represents the lower common denominator of documents across systems and software. Everyone with access to the Internet, even users with low speed modems and simple text based interfaces, can search, identify, browse, and download ASCII files containing relevant EMAP information.

Publication of longer documents containing tables and graphics requires additional consideration. Most documents are likely to be prepared using WordPerfect software since that is the standard word processing software being used by the Agency; however, not all users will have access to WordPerfect viewers. It is recommended that these documents be provided to users in three formats:

- As a WordPerfect 5.1 file, since WordPerfect is an Agency standard
- As unformatted ASCII text files to allow browsing by all types of users prior to downloading
- As portable document format (PDF) files to enable users to download, view, and print documents formatted as the authors intended the document to be published.

Additionally, it is recommended that longer documents be split into files representing major sections or chapters. Long documents are painful to browse or search on a computer screen and must be downloaded before they can be easily read. Users typically find it easier to browse shorter files to find relevant information before deciding to initiate downloading an entire document.

Publication of unformatted ASCII text files is recommended because not all Internet users have the capability of browsing WordPerfect formatted files online. By providing the ASCII text files, users can scan sections and decide if the full document is of use and is worth downloading. A disadvantage is that graphics and tables are not typically included with the ASCII text file; however, text alone is normally all that is needed to assess the utility of a document.

It is probably not necessary to provide ASCII text files for all chapters of long and complicated documents. ASCII text files for the Executive Summary or Abstract sections are typically all that are needed; however, the decision of what sections to provide as separate ASCII files will be left to the scientists and information management staff of the individual EMAP Task Groups.

In addition to ASCII and WordPerfect formats, it is recommended that documents containing graphics and tables be provided to users in portable document format (PDF) also. Using PDF reader software, users can locally view and print documents as they were meant to be seen by the authors of those documents. Until the Agency adopts a standard, use of Adobe Acrobat PDF is recommended. Support for the Adobe Acrobat format as a standard is likely to increase because of recent agreements between IBM, Netscape Communications, Inc., and Adobe Systems Inc. IBM plans to incorporate Acrobat Reader software as a standard feature on commercial PCs and Netscape is adding the software as a component of the browser software included with the corporation's World Wide Web client software. Apple Computer Inc. already includes Acrobat Reader software on some Macintosh computers. Adobe Acrobat Reader software is available through the Internet as freeware (http://www.adobe.com/software.html) and pointers to this software have been included in the directories of the public access server.

Task groups that would like to publish documents on the public access server can provide ASCII text files or WordPerfect files, or both. The EMAP IM Technical Support Staff will assist with converting WordPerfect files to ASCII text files and will provide PDF files for any documents containing graphics and tables. IM staff will also assist task groups with the preparation of HTML documents corresponding to the text files for executive summaries or abstracts. Guidelines for preparing ASCII formatted text files follow.

It is suggested that both hardcopy and electronic files be provided for large documents that EMAP Task Groups want to publish on the public access server. Technical Support Staff can then compare the electronic files to the hardcopy and check for formatting problems and potentially missing figures and tables. Task group staff should note those graphics that have not been included in the electronic files and were included in the hardcopy as "cut-and-paste" figures. To include these figures as part of the documents published on the public access

server, the figures will need to be electronically scanned and incorporated into the electronic files (See Section 5.4).

Note that EMAP as a whole will also maintain an internal server for documents of all types. Documents not approved for public release may still be made available to EMAP users via this server. WAIS search capabilities will be provided on this internal server. Should the details of preparation and submission of documents to this internal server differ significantly from those given above for the public access server, task groups will be so informed by separate communique.

5.1.1 Preparation of ASCII Files

Before converting a word processor document to text, convert to a non-proportional font (e.g., Courier) and set the page size so that no more than 70 characters will appear on a line. Eliminate margins unless they are specifically included (typed spaces) in the text and include a carriage return at the end of each line (this is usually an option in the "save as" instruction; the returns do not have to be hand entered). Remember that tabs (if they are not converted to spaces by the word processor save) may have different spacings on different machines or viewing software, and that special characters of any kind will probably not translate correctly.

5.2 METADATA

A requirement for publishing data sets using the public access server is that every data set is accompanied by a file providing documentation (metadata) for that data set. The metadata file is prepared as an ASCII text file following the procedures outlined above. The contents of the data set documentation file reflects the contents of the EMAP data catalog (Strebel and Frithsen 1995) and is an integral part of the EMAP Information Management System. Information in the catalog is stored within the relational tables of the EMAP distributed data base and is provided to users as customized views of the data base. Catalog information extracted from the data base to create data documentation files represent a snapshot of the data base at the time files are created.

The data set catalog and, therefore, data documentation files, contain the following sections:

1. Data set identification
2. Investigator information
3. Data set abstract
4. Objectives and introduction
5. Data acquisition and processing methods
6. Data manipulations
7. Data description
8. Geographic and spatial information

9. Quality control and quality assurance
10. Data access and distribution
11. References
12. Table of Acronyms

This information provides potential users of the data with the information necessary to evaluate if the data can be used for purposes other than those envisioned by the designers of EMAP. Complete descriptions of the catalog along with examples of catalog entries are provided in the document prepared to assist task groups with the completion of catalog entries for EMAP data (Strebel and Frithsen 1995).

5.3 DATA

Data may be submitted as ASCII text files, or in a variety of binary formats. ASCII, comma delimited values (with headers) are the least common denominator. Other formats are for the convenience of particular groups of users. Header information should include, at the least: data set name, file name and creation/revision date, number of header lines, number of data lines; missing value indicator; variable names and definitions, one per line; comma separated data values - one set per line, with all variables listed. An example of an ASCII data file is provided in Table 5-1.

Obviously, some data can be provided only in binary format. Such binary files might include spreadsheet files (e.g., Excel), statistical analysis files (e.g., SAS), files formatted for geographic information systems (e.g., ARC/Info), and other types of binary and image data.

5.4 IMAGES, VIDEOS, SOUND

Binary files containing images, video, and sound can be supported by many WWW browsers, as well as downloaded for local use. For the most part, the file formats are application specific, although TIFF is the most commonly supported image format across multiple platforms. For images to be used with WWW pages, GIF provides a good compressed format (although there are some unresolved proprietary issues with this copyrighted format and associated software). If lossy compression is acceptable, (i.e., the reconstructed image will not have all of the detail and resolution of the original), the JPEG format may be used.

No specific recommendations for video and sound files are made at this time; however, task groups are encouraged to work with the technical support staff to experiment with including this kind of information with data set documentation and files published using the public access server.

Table 5-1. Example of an ASCII data file with recommended format for header information

```
Data set name:  Diel values for dissolved inorganic carbon
File name:  DISINCAR.DAT
Created: March 20, 1995
Number of header lines: 17
Number of data lines:  24
Missing value indicator: '.'
Variable names and description:
        OBS             Observation number (numeric)
        DATE            Sampling date (YYDDD)
        TANK            Experimental tank (numeric)
        DAYT            Number of daylight hours (numeric)
        REP             Replicate number (numeric)
        CO2DAWN1        CO2 concentration at dawn on day 1 (mg/liter)
        CO2DUSK1        CO2 concentration at dusk on day 1 (mg/liter)
        CO2DAWN2        CO2 concentration at dawn on day 2 (mg/liter)
        CO2DUSK2        CO2 concentration at dusk on day 2 (mg/liter)


1,88159,4,15.08,1,23.4824,21.6127,22.8783,.
2,88159,4,15.08,2,23.4532,.,
        22.7631,.
3,88159,4,15.08,3,.,
        21.9228,.,      .
4,88159,5,15.08,1,22.9308,22.7657,.,.
5,88159,5,15.08,2,23.4930,22.8233,22.4554,.
6,88159,5,15.08,3,23.4055,22.7605,22.5387,.
7,88159,6,15.08,1,22.5986,.,
        22.6247,.
8,88159,6,15.08,2,22.5178,21.2158,22.5204,.
9,88159,6,15.08,3,.,
        21.5846,22.5751,.
10,88159,7,15.08,1,22.9754,22.7422,22.3982,.
11,88159,7,15.08,2,22.6247,22.5230,22.3411,.
12,88159,7,15.08,3,22.7422,22.6012,22.3178,.
13,88159,8,15.08,1,21.8688,21.7483,21.5693,.
14,88159,8,15.08,2,21.9845,22.4190,22.3697,.
15,88159,8,15.08,3,21.9974,22.1677,21.9408,.
16,88159,9,15.08,1,22.2738,22.3152,22.4060,.
17,88159,9,15.08,2,22.1522,22.0258,22.1858,.
18,88159,9,15.08,3,21.5999,22.1418,.,.
19,88159,12,15.08,1,22.6560,22.3437,22.3023,.
20,88159,12,15.08,2,22.4658,22.2272,22.0799,.
21,88159,12,15.08,3,22.4528,21.9125,22.2919,.
22,88159,14,15.08,1,22.5204,22.0129,22.0206,.
23,88159,14,15.08,2,22.7971,22.0129,22.4918,.
24,88159,14,15.08,3,22.1496,22.0902,22.0026,.
```

5.5 HTML FILES

The TSS will prepare the WWW interface and many documents for the server using Hypertext Mark-Up Language (HTML). This allows formatting instructions and hyper-links to be encoded in simple ASCII text files. If Task Groups have prepared HTML versions of documents, they can be used by the TSS if they conform to the general standards being used on the public access server. Consult with the TSS about this when submitting your material. (Note: All interface HTMLs will be maintained by the TSS, and substitute interface HTMLs should not be created by the Task Groups. Suggestions for interface development, however, are welcome at any time).

IM staff will assist task groups with creating HTML documents. HTML is likely to be most effective if applied to the Executive Summary and Abstract sections of reports and technical papers. At this time, it is not recommended as a general approach that entire documents be converted to HTML because of the resources needed to produce good quality HTML documents.

5.6 DATA COMPRESSION

It is recommended that task groups do not compress files prior to submission. Where appropriate, the TSS will create compressed files using commonly software for PC platforms (PKZIP) and UNIX environments (TAR). It is likely that compression will be used for complex documents containing embedded graphics and tables.

The creation of PDF files typically produces large files. To some extent, the size of these files can be controlled by not embedding fonts within the PDF files and by specifying lower resolution for graphics. The compression of PDF files may introduce incompatibilities with the viewers that are being incorporated into the next versions of WWW client software.

6.0 GENERAL MAINTENANCE

The Technical Support Staff (TSS) will provide general maintenance of the public access server, including the EMAP data and information posted on it. Several areas of importance will be: data submission tracking systems, backups, security, user interface access statistics, and user comments.

6.1 TRACKING SYSTEM

The data submission procedure requires, and will be made more efficient by, a tracking database. The database provides an inventory of the files intended for publication using the public access server. This could be a part of the IMS ORACLE RDBMS (which would allow contact links, etc.). Primary information to be recorded includes: submission form information, approval check completion dates, technical review completion date, scientific review completion date, limited access server posting date, public release date, along with a tracking record identifier, comments, and revision date set pointers and deletion date and reason. The TSS will maintain the data submissions tracking system. Examples of the type of information that should be included in the data base are provided in Table 6-1.

6.2 BACKUPS

Routine backups should be performed by the Technical Support Staff to provide recovery from hardware failures, software glitches, and natural disasters. At specified intervals, some of these backups should be archived to maintain a set of historical snapshots of the EMAP data posted on line.

6.3 SECURITY

All Agency security procedures will be followed and enforced by the TSS. Task groups will be notified of security violations and/or attacks affecting their data and information. Since the public access server will be outside of the security firewall, it is a possible target for malicious activity. Task groups should periodically make spot inspections of the data and information on the server, and report suspected modifications and/or misuse to the TSS.

6.4 USER INTERFACES

WWW pages, Gopher menus, and WAIS indexes that provide user access to the data will be prepared and maintained by the TSS. Suggestions for interface changes in updates, particularly with respect to task group material, are encouraged.

Table 6-1. Information to be included in the inventory tracking database.	
File inventory information	Inventory tracking identification Date received Version number
File information	File name File type File size File date
Task group information	Task group Submitted by
Approval information	Approved by Approval date Approved for limited access server Approved for public access server Cleared for release after Expiration date Restrictions on dissemination
Review information	Technical reviewer Date technical review completed Scientific reviewer Date scientific review completed
Publication information	Date of task group final approval Date moved to limited access server Date moved to public access server
Revision information	Name of revised file Date of revision Archival status of file

6.5 ACCESS STATISTICS

A record of access to the data and information should be maintained and regularly summarized by the TSS. This record can be used to establish which files are the most "valuable" to users, identify the user clientele, and provide other feedback for optimizing the system for productive use (e.g., what tools and indexes are the users using most, are the directories logically organized for the searches being conducted, should some data sets be combined, should some be removed, etc.?)

6.6 COMMENTS FROM USERS

Online feedback mechanisms will be provided to capture user comments and/or provide avenues for them to supply feedback. The TSS will distribute any such feedback to the SRS, the IMTC, and appropriate task group personnel.

Comments from users of the EMAP data and information on the Agency's public access server are useful. Users often use the data in unanticipated ways, revealing both its strengths and its weaknesses. An ideal situation would be to capture user feedback in future revisions and documentation updates for a data set.

User comments about access mechanisms and organization, and other similar technical issues, will be resolved by the IMTC, the TSS, and the SRS as appropriate. All task group users should be active participants in providing this feedback, and where possible, soliciting it from their collaborators who use the public access server to obtain EMAP data.

7.0 SUPPORT TO TASK GROUPS FROM THE EMAP IM TEAM

Publishing EMAP data using a public access server will be a cooperative activity. While the Task Groups will be responsible for the contents of their sections on the server, they cannot be expected to maintain the common infrastructure or provide the resources for significant additional work entailed by publishing their data and information in this way. The EMAP Information Management staff has the responsibility to provide appropriate support to the Task Groups in this endeavor, including organizational and procedural guidelines, maintenance, physical and logical security (e.g., firewalls, ftp logs), and backups and historical archives.

Developing common guidelines and other required infrastructure (e.g., the data submission tracking system) will be an activity of the EMAP IM staff. Specific assistance will also be provided for data reformatting and pre-publication review. Indexing the submitted data and developing access tool scripts (e.g., html files) will also be a responsibility of the IM staff.

These responsibilities will require both technical and scientific support to be provided by the IM staff. In general, these functions will be separated into an on-site technical support component staff and an off-site scientific review component. Where issues of scientific evaluation or judgement may arise, the scientific review staff will be the primary resource. Routine technical activities, especially hardware and software related activities, will fall to the technical support staff.

7.1 DOCUMENT SUBMISSION SUPPORT

Format guidelines are provided in Section 5.0. Technical support and scientific review staff will assist EMAP Task Groups with the preparation of documents for publishing using the public access server. This support will include: preparation of ASCII formatted files from WordPerfect files; preparation of PDF files from WordPerfect and graphics files; and scanning of graphics into electronic files. Formatting data conversions will be provided under the supervision and review of the SRS to insure that all quality issues are satisfied and no scientific errors are introduced.

7.2 DATA SUBMISSION SUPPORT

Format guidelines are provided in Section 5.0. Organizational guidelines are provided in section 3. One of the responsibilities of the SRS will be to suggest data organization that both assists task groups to publish their data and makes access to and integration of EMAP data friendly to the variety of users expected.

8.0 SUPPORT TO USERS

Principal support for the users of the USEPA's public access server is provided through existing mechanisms by staff supported by the Agency's National Data Processing Division in Research Triangle Park, NC. This staff maintains the information on the server and provides users with information about Internet information discovery and access software (World Wide Web, Gopher, WAIS), and other software (PDF viewers, for example) that assists users with use of the data and information provided on the public access server.

EMAP technical support staff will provide similar support to the principal users of EMAP data. Support may include assistance with the installation of Internet software tools and the use of these tools to identify and obtain EMAP data sets of interest. The EMAP technical support staff will also provide links to other servers containing environmental data that are potentially of use to EMAP users. Additional user support will be provided by EMAP technical support staff through continued maintenance of the "What's New" file containing notice of data and information added to the EMAP portion oft he server. The EMAP IM technical support staff will work with the staff maintaining the public access server to ensure that WWW and Gopher servers are registered with appropriate indexing agencies, thus improving user access.

An important user support function provided by the EMAP technical support staff and the scientific review staff is to evaluate and respond to comments and questions provided by users of the EMAP portion of the public access server. To ensure continued use of the system, it is important that users receive a reply to all questions by the appropriate staff within the program.

9.0 LITERATURE CITED

Bills, H.M. 1994. Memorandum to ORD Laboratory Directors. December 5, 1994. U.S. Environmental Protection Agency, Washington DC.

Franson. 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.

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

Hayes, B. 1994. The World Wide Web. *American Scientist* 82: 416-420.

Meeson, B.W., D.E. Strebel, and D.R. Landis. 1993. Producing a CD-ROM: A Workbook. April 1993. Goddard Distributed Active Archive Center, NASA Goddard Space flight Center, Greenbelt, MD.

Kirkland, L.L. 1994. EMAP Quality Management Plan. U.S. Environmental Protection Agency, Washington DC.

NRC. 1994. Review of EPA's Environmental Monitoring and Assessment Program: Forest and Estuaries Components. Pre-Publication Draft. National Research Council, Washington DC.

Schatz, B.R. and J.B. Hardin. 1994. NCSA Mosaic and the World Wide Web: Global hypermedia protocols for the Internet. *Science* 265: 895-901.

Shepanek, R. 1994. EMAP Information Management Strategic Plan: 1993-1997. EPA/620/R-94/012. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP). Washington DC.

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

USEPA. 1994. EMAP Information Management Virtual Repository. Draft September 9, 1994. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Washington, DC.

USGAO. 1993. EPA's plans to improve longstanding information resources management problems. GAO/AIMO-93-8. September 1993. U.S. General Accounting Office, Washington DC.

Waugh, M. and M. Weaver. 1994. Data submission procedures for EPA Internet Servers. Document #698/001. December 21, 1994. U.S. Environmental Protection Agency, National Data Processing Department, Research Triangle Park, NC.