SCIENTIFIC DOCUMENTATION FOR EMAP DATA: GUIDELINES FOR THE INFORMATION MANAGEMENT CATALOG

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 Road
Columbia, Maryland 21045

30 April 1995

The suggested citation for this report is:

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 U.S. Environmental Protection Agency, Office of Modeling, Monitoring Systems and Quality Assurance, Washington, DC. Prepared by Versar, Inc., Columbia, MD.

EXECUTIVE SUMMARY

The Environmental Monitoring and Assessment Program (EMAP) is an interagency effort coordinated by the U.S. Environmental Protection Agency and designed to collect information to assess the condition of the nation's ecological resources. The Information Management System for EMAP was developed to capture, preserve, and provide to users data and information collected and prepared by the program. The report describes the data catalog component of the EMAP Information Management System. Together with the data directory and dictionary, the catalog is one of three components that provides information about data (metadata) to users. The catalog contains the detailed, scientific documentation about data that enables assessment scientists and managers to understand the conditions, assumptions, and methods under which data were collected and compiled.

Presented in this report is a summary of the metadata components for the EMAP Information Management System, design requirements for the catalog, and a description of the catalog structure that was developed in response to those requirements. Requirements for the catalog were determined by EMAP assessment scientists and other users through multiple joint application design sessions and feedback provided through EMAP task group information managers. These requirements were refined based upon the overall vision for the EMAP Information Management provided by the Strategic Plan (Shepanek 1994) and review of existing standards and procedures for the completion of scientific documentation. In response to these requirements, the catalog was designed as an integral component of the EMAP relational data base. The structure of the catalog is, therefore, a set of fields organized in relational tables.

The drafting of scientific documentation is presented as a collaborative effort between scientific investigators and information management staff. Guidelines presented in this report for scientific investigators present writing a catalog entry as an analogous process to writing a scientific publication. Guidelines presented for information management staff focus on formats and the definition of specific fields. This report also contains examples of data documentation to be used by others for the compilation of additional catalog entries.

ACKNOWLEDGMENTS

This report was developed with input from Paul Cole, Gary Collins, Steve Hale, Melissa Hughes, Judith Lear, Jane Lovelace, Doug Mann, Jeff Rosen, and David Strevel. Technical oversight and review as provided by Robert Shepanek. This report was prepared in response to work assignment 4-12 of contract 68-DO-0093 from the U.S. Environmental Protection Agency to Versar, Inc.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

ABBREVIATIONS

ASCII	American Standard Coding for Information Interchange
BOREAS	Boreal Ecosystem-Atmosphere Study
EMAP	Environmental Monitoring and Assessment Program
EOSDIS	Earth Observing System Data and Information System
FGDC	Federal Geographic Data Committee
FIFE	First International Satellite Land Surface Climatology Project Field Experiment
FTP	File Transfer Protocol
IM	Information Management
IMS	Information Management System
JAD	Joint Application Design
LTER	Long-Term Ecological Research
NASA	National Aeronautics and Space Administration
NRC	National Research Council
NSF	National Science Program
QA/QC	Quality Assurance/Quality Control
RDBMS	Relational Data Base Management System
SOP	Standard Operating Procedure
TGIM	Task Group Information Manager
TIFF	Tagged Image File Format
URL	Univesal Resource Locator
USEPA	U.S. Environmental Protection Agency
UT	Univesal Time
WAIS	Wide-Area Information Server
WWW	World Wide Web

1.0 INTRODUCTION

The U.S. Environmental Protection Agency (USEPA) is coordinating the Environmental Monitoring and Assessment Program (EMAP), a multiagency effort to establish a national monitoring program designed to collect the information necessary to assess the condition of the nation's ecological resources. The purpose of this document is to provide scientists and information managers within the program guidance concerning the writing of detailed documentation for EMAP and other environmental data sets. Detailed documentation for data enable assessment scientists and managers to understand the conditions, assumptions, and methods under which data were collected and compiled, thus allowing the data to be used for some purpose. The detailed documentation is one of several levels of information about data (metadata) included in the EMAP Information Management System.

1.1 EMAP INFORMATION MANAGEMENT SYSTEM

The EMAP Information Management System (IMS) was designed based upon evaluations of the diverse types of users the system must serve; the great diversity of data types that must be captured, stored, and provided to those users; the types of environmental assessments that those users are likely to complete; and, the analytical, statistical, and data visualization tools users will need to complete those assessments. The results of those evaluations formed the basis for the EMAP Information Management Strategic Plan (Shepanek 1994), a document that has guided the activities of the EMAP Information Management (IM) Task Group.

The core of the EMAP IMS is a distributed relational data base containing both data and descriptive information (metadata) about data residing in the data base or in files external to the data base. Metadata are a central component of the IMS. A fundamental requirement outlined in the strategic plan is that metadata components must be both flexible and robust to meet the needs of EMAP users. When fully implemented, EMAP data will be collected throughout the United States and involve the resources of multiple federal agencies. Initial users of EMAP data represent resource and coordinating groups within the program. Each of these groups has developed their own information management centers and some groups have implemented multiple data centers based upon the regional implementation of monitoring efforts. In addition to these initial data users, a diverse set of users outside of the program has arisen. These users represent various government agencies, industry, academia, and private nongovernment organizations.

1.2 METADATA COMPONENTS

To meet the data needs of these varied users, EMAP metadata components are structured and organized so that users can easily find the information needed to select EMAP data sets, while not becoming inundated with unnecessary information. This organization is based upon identifying components of metadata that may be logically grouped based upon the expectations of data users.

Metadata components are often referred to using terms that are not uniformly defined or applied by the information management community (Strebel and Frithsen 1991). Terms such as inventory, directory, catalog, and dictionary refer to different levels of metadata, the elements of which are typically defined by individual programs and not by generally accepted guidelines.

The three principal metadata components being developed as part of the EMAP IMS are the data set directory, catalog, and dictionary. Each component provides users with different types of information needed to identify, describe, and locate data sets. The definitions recommended for EMAP for each of these metadata components (Table 1-1) are consistent with those given in the NASA Earth Sciences Lexicon (NASA 1991), and the outline for the EMAP virtual repository presented in the EMAP Information Management Strategic Plan (Shepanek 1994; USEPA 1994).

Metadata components are designed to meet the needs of data base personnel and scientists and managers that use the data. The functionalities provided by the data set directory, catalog, and dictionary for data base personnel and other users are different (Table 1-2). In general, data base personnel use metadata to index, track, and organize data; others use metadata to identify what data are available and to obtain information used to understand the data. The functions are complementary, but in general the documentation required by data base personnel should not obfuscate other users' understanding of the data.

In a previous report (Strebel and Frithsen 1991), metadata components were outlined and organized in a manner analogous to a scientific publication. This analogy has since been further developed (Strebel and Meeson 1992; Strebel et al. 1994a) to reflect the importance of metadata development to the general scientific community.

Completely describing a data set is analogous to writing a manuscript for publication in a scientific journal (Strebel et al. 1994a). The metadata analogy to the manuscript is the data set catalog, which is also referred to as detailed documentation. This detailed documentation includes information concerning the originators of the data set, the general purpose for which the data were collected, sampling and laboratory methods, descriptions of the data and any manipulations or transformations of the data, related quality control and quality assurance measurements, procedures necessary for data access, and references to publications that use the data set.

Table 1-1. Working definitions for EMAP metadata components

Metadata Component	Definition
Data set directory	Summarized data set documentation. A uniform set of descriptions of a large number of data sets, containing information suitable for making an initial determination of the existence and nature of each data set (NASA 1991).
Data set catalog	Detailed data set documentation - also referred to as the scientific documentation. A uniform set of detailed descriptions of a number of data sets and related entities, containing information suitable for making a determination of the nature of each data set and its potential usefulness for a specific application (NASA 1991).
Data dictionary	Fundamental data set documentation. The data dictionary provides a short scientific description of a parameter or variable in a data set, along with format and other basic information used in storing, searching, and displaying the data item. The dictionary contains information about the contents of each table in the relational data base.

Table 1-2. Functions of metadata components

Metadata Component	Use by Data Base Personnel	Use of Others
Directory	- Index and track data	- Identify data sets - Select data sets of interest
Catalog	- Record ancillary information about data	- Obtain descriptions of the data
Dictionary	- Organize the data - Define data formats	- Select data items of interest - Understand how to use the data

A summary of the detailed documentation is provided in the data set directory. Directory entries are analogous to the abstract of a scientific paper and contain information to assist the data user in identifying data sets that may be of interest. The directory is linked to the data set catalog so that users may quickly locate additional information concerning data sets of interest. Further, directory level information helps data management personnel index and track data sets. Details for the design of the EMAP data directory and provided in Frithsen and Strebel (1995).

A section of a data set catalog is directly linked to the data dictionary and provides technical specifications for each data item. This fundamental documentation is used by data management personnel to organize the data and by data users to select data items of interest and to determine output formats.

In addition to summarized, detailed, and fundamental documentation (directory, catalog, and dictionary metadata components, respectively), auxiliary documentation may be used to store additional information related to data sets. This auxiliary documentation can include methods manuals, photographic and electronic images of field and laboratory data sheets and quality assurance audit reports, and publications that use the data set in question. The method of access to this type of information generally depends upon user defined requirements. The EMAP Strategic Plan (Shepanek 1994) outlines a design for a virtual repository (USEPA 1994) linking common metadata components (directory, catalog, and dictionary) with auxiliary documentation in related data bases, thus providing users with links between related information about EMAP and EMAP data sets.

The term inventory was used in earlier metadata discussions to refer to a subset of the information contained within the data set directory. In this context, the inventory contains information used to index and track data sets only, with little summary of the contents of the data set. This definition of inventory is redundant with what is being referred to in the directory, and is in conflict with the definition embraced by other environmental science programs (NASA 1991).

The term inventory should refer to a uniform set of descriptions of elements, or granules of a data set. (A data set granule is the smallest aggregation of data that is independently managed.) The descriptions contained within the inventory provide the information needed to select the data granule of interest. Granule descriptions typically include temporal and spatial coverage, data quality indicators, and physical storage information. The contents of the inventory, therefore, should be tailored to the set of data granules to be described. Guidance on the contents of specialized inventories will be provided in a future report.

Principal components of the directory and catalog are shown in Figure 1-1. Also shown is how the components relate to the processes affecting the maturation of data from collection to publication. The development of metadata is presented as a process that begins with compilation of objectives and collection methods for the catalog by sample collection investigators. When the data are electronically captured in the IMS, IM staff begin a directory entry. Additional details are provided for the directory as the data are

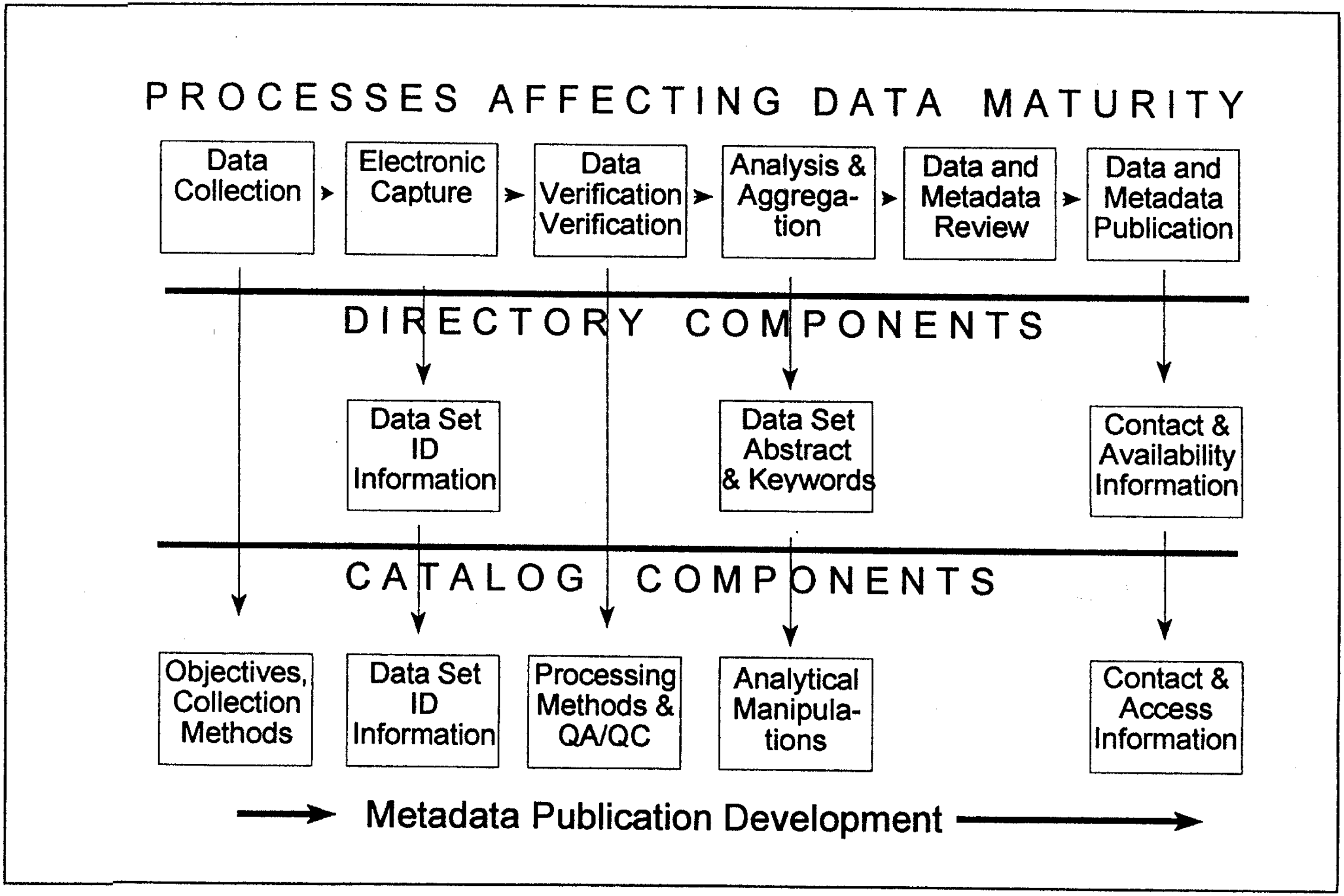
PROCESSES AFFECTING DATA MATURITY
Data Collection
Electronic Capture
Data Verification Verification
Analysis & Aggregation
Data and Metadata Review
Data and Metadata Publication
DIRECTORY COMPONENTS
Data Set ID Information
Data Set Abstract & Keywords
Contact & Availability Information
CATALOG COMPONENTS
Objectives, Collection Methods
Data Set ID Information
Processing Methods & QA/QC
Analytical Manipulations
Contact & Access Information
Metadata Publication Development

Figure 1-1. Principal components of the directory and catalog as they relate to the maturation of data

processed and aggregated and various quality control information is assessed. Both the directory and the catalog undergo a review by internal and external environmental and information management scientists to assess completeness and usefulness to future assessment scientists and interested public users.

1.3 ORGANIZATION OF REPORT

This section of the report has presented background information for the development of the EMAP IMS and the metadata components included in the EMAP IMS. The user defined requirements for the detailed, scientific documentation component of metadata (the data catalog) are presented in Section 2 along with a summary of the individual fields that comprise the catalog. Section 3 provides guidance to the EMAP assessment scientist for the writing of a catalog entry. Examples of catalog entries are provided in Section 4 to show how information is formatted within specific fields. Guidelines for the display of catalog information to users is provided in Section 5. A detailed appendix is provided to define for information management staff the format and suggested content of each field included in the catalog.

2.0 DATA CATALOG REQUIREMENTS AND CONTENTS

Specific requirements for the design and composition of the data catalog were developed based upon the vision provided in the EMAP Information Management Strategic Plan (Shepanek 1994), input from representative users of EMAP data, insight gained from data documentation efforts completed for other programs, and general guidance from the scientific information management community. Input from representative users of EMAP data was obtained through various joint application design (JAD) sessions (USEPA 1993a,b; Palmer and Fields 1994). Additional input was provided by task group information managers (TGIM) who interact with EMAP assessment staff almost daily. Additional requirements for the catalog were developed from an evaluation of the lessons learned from the development of data documentation for other projects, specifically NASA'S EOSDIS, FIFE, and BOREAS projects and the National Science Program's (NSF) Long-Term Ecological Research (LTER) Program (Michener et al. 1990; Meeson et al. 1993; EOSDIS 1994; Strebel et al. 1994b; Justice et al. 1995). Requirements for the EMAP data catalog also reflect guidance provided from various interagency efforts concerning information management for environmental data. These efforts include the National Research Council's report concerning data management for global change assessment (NRC 1991) and the Federal Geographic Data Committee's guidance for the compilation of spatial metadata (FGDC 1994).

2.1 REQUIREMENTS

The following general requirements have been defined for catalog documents:

! The document must be a stand-alone description of the data set

The detailed documentation presented in the data set catalog should contain enough information so that a potential user can fully evaluate the application of a data set for a particular purpose. This requirement is a component of what has been termed the "20-year rule" for data documentation (NRC 1991), i.e.: "Will someone 20 years from now, not familiar with the data or how they were obtained, be able to find data sets of interest and then fully understand and use the data solely with the aid of the documentation archived with the data set?"

! The document must be structured for usability

Users should be able to browse the detailed information in the catalog starting at specific sections of interest (for example, methods, data descriptions, publications, etc.). They should also be able to read the entire document as if it were a scientific paper describing the data set.

! Catalog information must be linked to the data

Access to the information in the data set catalog should also be provided directly from the EMAP data base. This represents a data management challenge because documentation about data is organized around data sets but the data are stored in a combination of relational tables in a data base and conventional data files in ASCII, SAS, or other formats.

! Detailed documentation must accompany distributed data

A stated requirement of EMAP (Kirkland 1994) is that complete data documentation will accompany data distributed by the program. In this way, users will be provided information with which to understand the appropriate uses and limitation of the data. The documentation that needs to accompany data is contained within the data set catalog.

! The design of the catalog must minimize redundancy.

Complete scientific documentation includes information that may be stored in other components of the EMAP IMS. Information for personnel contacts, methods, and publications are examples of catalog information that is already a part of the overall EMAP relational database. The catalog needs to be organized so that this information is stored once, but accessed in various forms. This is consistent with the concept of a virtual repository for information about the program, EMAP data sets, and the EMAP IMS (USEPA 1994).

! Catalog information must be linked to the EMAP Data Set Directory

The data set directory (Frithsen and Strebel 1995) will, in most cases, be the first stop for users interested in obtaining information about EMAP data sets. The directory should be linked to the data set catalog so that potential users can obtain detailed information about data sets identified through the directory and thus, select data sets of interest.

! The data set catalog must be linked to the data base dictionary

Descriptions of the data that are included as part of the data set catalog are derived directly from the data base dictionary; therefore, the catalog and dictionary need to be linked to ensure that changes within the data base are represented in the detailed documentation.

2.2 DATA CATALOG COMPONENTS

There is a logical organization for the detailed information contained within a data set catalog entry. This organization reflects the expectations of a potential user who might browse the information presented in the catalog. The organization of the data catalog is patterned after that of a scientific paper (Strebel et al. 1994a). Title and author information is followed by an introduction presenting the objective and purpose of the data set. There is a methods section describing how the data were collected or acquired and the analytical procedures used to process samples. A separate section describes any analytical transformations used to prepare the data in the data set.

Similar to the scientific paper, the introduction and method sections are followed by descriptions of results. The results presented in the catalog present a description of the contents of the data set and the geographic and spatial scope represented in the data. Summarized results of quality control and quality assurance procedures are also presented.

The data set catalog ends with a discussion of how to access the data and other means by which the program has distributed the data. This section is followed by a list of bibliographic references pertaining to the collection and processing of the data, and demonstrating how the data has been used.

The information presented in the data catalog supplements information provided in the data set directory and the data base dictionary; however, selected information included in the directory and dictionary are also included in the data catalog. The design of the catalog is such that the information stored in the three main metadata components (directory, catalog, and dictionary) is brought together in various views that in turn make the data catalog.

The components of the data set catalog are organized into 13 sections representing separate groups of related information. The 13 sections are listed in Table 2-1.

Each section includes fields used to organize and store information contained within the catalog. These fields generally contain text which in some cases may be several paragraphs to pages in length. The fields have been defined to assist information managers responsible for the completion of data catalog entries. By providing specifically defined fields instead of fewer, general fields, the information manager and the scientist that produced the data set (the catalog authors), are prompted for the information necessary to complete the data set catalog entry.

Table 2-1. Major sections of the EMAP Data Catalog

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. Glossary and Table of Acronyms
13. Personnel Information

The specific fields defined for each of the 13 section of the catalog are summarized in Table 2-2. Several fields included in the catalog are replicated from the data objects that represent groups of tables describing the data directory, contacts, methods, and publications. These fields are noted in Table 2-2. The components of the tables describing the data directory are described in the guidance document for the data directory (Frithsen and Strebel 1995). Earlier drafts of this document were the basis for the data object describing contacts within the program. The directory and the contact data objects are already implemented in the EMAP Oracle data base. The tables describing the methods data object are being designed by the Methods Coordinating Group with assistance from the EMAP Information Management User Interaction and Planning Team. The components of the tables describing the EMAP publications data object are defined in this document and were developed based upon the information presented in Lear and Chapman (1994) and the components of the data base used to develop that publication (Lear, personnel communication).

Table 2-2. Individual fields included in the EMAP Data Catalog. Fields replicated from other data objects in the EMAP relation data base are footnoted. Numbers refer to major sections defined in Table 2-1.

1. Data Set Identification Title of Catalog document Author(s) of the Catalog entry Catalog revision date Data set name[a] Task Group[a] Data set identification code[a] Version[a] Requested Acknowledgement	2. Investigator Information Principal Investigator Sample Collection Investigator Sample Processing Investigator Data Analysis Investigator Additional Investigator
3. Data Set Abstract Abstract of the Data Set[a] Keywords for the Data Set[a]	4. Objectives and Introduction Program Objective Data Set Objective Data Set Background Information Summary of data set parameters
5.1 Data Acquisition Sampling Objective Sample Collection Methods Summary Beginning Sampling Date Ending Sampling Date Sampling Platform Sampling Equipment Manufacturer of Sampling Equipment Key Variables Sampling Method Calibration Sample Collection Quality Control Sample Collection Method Reference Sample Collection Method Deviations	5.2 Data Preparation and Sample Processing Data Preparation Objective Data Processing Methods Summary Sampling Processing Method Calibration Sample Processing Quality Control Sample Processing Method Reference Sample Processing Method Deviations
6. Data Manipulations Name of New or Modified Value Data Manipulation Description Data Manipulation Examples Data Manipulation Computer Code File Data Manipulation Computer Code Language Data Manipulation Computer Code	7.1 Description of Parameters Parameter Name SAS Parameter Name Parameter label or description Units of measurement Parameter data type Precision to which values are reported Accuracy of the data values Minimum Value in Data Set Maximum Value in Data Set
7.2 Data Record Example Column Names for Example Records Example Data Records	7.3 Related Data Sets Related Data Set Name Related Data Set Identification Code

Table 2-2. (Continued)

8. Geographic and Spatial Information Minimum Longitude Maximum Longitude Maximum Latitude Minimum Latitude Name of the area or region Direct Spatial Reference Method Horizontal Coordinate System Used Resolution of Horizontal Coordinates Units for Horizontal Coordinates Vertical Coordinate System Resolution of Vertical Coordinates Units for Vertical Coordinates	10. Data Access Data Access Procedures Data Access Restrictions Data Access Contact Person Data Set Format Information Concerning Anonymous FTP Information Concerning Gopher Information Concerning World Wide Web EMAP CD-ROM Containing the Data set
11. References Reference Type[d] Reference Author[d] Reference Author's Affiliation[d] Title of Reference[d] Journal or Volume Title[d] Journal or Volume Editor[d] Page and Volume Reference[d] Date the Reference was Published[d] Location of Publishing Organization[d] Name of Publishing Organization[d] Reference Report Number or Other ID[d] Procite Record Number for the Reference[d]	9. Quality Control/Quality Assurance Measurement Quality Objectives Quality Assurance/Control Methods Actual Measurement Quality Sources of Error Known Problems with the Data Confidence Level/Accuracy Judgement Allowable Minimum Values Allowable Maximum Values QA Reference Data
12. Table of Acronyms Glossary Term or Acronym Definition of Glossary Term or Acronym	13. Personnel Information Formal Title[a,b] Last Name[a,b] First Name[a,b] Middle Initial[a,b] Role[a,b] Line 1 of Address[a,b] Line 2 of Address[a,b] Line 3 of Address[a,b] Line 4 of Address[a,b] City[a,b] State[a,b] Zip Code[a,b] County[a,b] Voice Phone Number[a,b] Fax Phone Number[a,b] Email Address[a,b] Email Network[a,b] Additional Email Information[a,b]

[a] Field included in the Data Set Directory
[b] Field included in the Contacts data base object.
[c] Field included in the Methods data base object.
[d] Field included in the Publications data base object

3.0 WRITING THE CATALOG ENTRY

The catalog represents a technical document written to convey technical information to a scientific audience. Catalog entries are best written by the investigators who directed the collection and analysis of the data being documented. In general, information management staff will not have the technical background to write or review catalog entries; however, the data documentation effort represents a collaborative effort between scientific and information management staff. Principal investigators provide technical descriptions that enhance understanding of the data while information management staff help to organize those descriptions within the information management system.

The data set catalog contains detailed, background information about a data set. This information includes descriptions of data collection methods, laboratory and analytical procedures, summaries of quality assurance results, and other types of information that can be utilized by potential users to understand and begin using the data for some purpose. These uses of the data may be entirely different from the uses for which the data were originally collected. Information contained within the catalog should be of sufficient detail and completeness to minimize the need for most users to access additional sources of information.

This chapter lays out the organizational structure and content of the catalog entry at a level appropriate for primary authors. The author of a catalog entry should normally be a scientist familiar with the data set who will approach this task as one of writing a scientific paper about the data set. In most cases, the primary author will be the investigator responsible for collecting and analyzing the data. Some technical details, of course, will be provided by the cognizant TGIM when the document is finalized. It is the responsibility of the TGIM to represent EMAP IM in obtaining, submitting, and maintaining catalog entries that adequately describe all EMAP data sets. The detailed technical reference for the material in this section is given in the Appendix.

3.1 DATA SET IDENTIFICATION

This section of the catalog contains information used to identify the data set (Table 3-1). The data set name gives a convenient user reference for the data set, while the data set identification allows the information management staff to build technical links between the catalog document and the data set itself.

Table 3-1. Catalog components describing data set identification information

Field	Field Description
Title:	Title of Catalog document
Cat Author:	Author(s) of the Catalog entry
Cat Rev Date:	Catalog revision date
Data Set Name:	Data set name
Task Group	Task Group Code
Data Set Id:	Data set identification code
Version	Version number
Req Acknowl	Requested acknowledgement

Discussion: These fields identify the document and the data set to which it refers. The title, author, and catalog revision date are used to identify and track the document itself. The title of the data set catalog entry, in most cases, will be similar to the name of the data set, although perhaps more succinct. Catalog author is one of the defined roles supported in the contacts data base object. The field identifies people who have contributed to writing the catalog entry and may be repeated as many times as is necessary to identify multiple catalog authors. Additional information about each named individual is provided in the Personnel Information section of the catalog (Section 3.13). The catalog revision date field is a mandatory field used to record when the catalog entry was last revised (e.g. to reflect the results of new analyses or to incorporate information concerning additional data uses, limitations, and publications). The data set identification code contains information about the EMAP task group from which the data set originates. Along with Task Group and Version Number, the Data Set Identification Number uniquely identifies a data set or view. Because the data set may be referenced in scientific publications, a suitable acknowledgement for its use should be suggested. Data sets collected by EMAP staff under programmatic auspices will have a general EMAP acknowledgement statement. Where a data set is the product of a specific investigator's work, and should be so referenced, the investigator should provide appropriate acknowledgement text.

3.2 INVESTIGATOR INFORMATION

This section of the data set catalog contains information identifying the individuals, or group of individuals, who produced the data set (Table 3-2). This normally means principal investigators and their immediate associates. Reference to whom should be contacted to obtain the data, and data acquisition procedures, are handled in a separate section (Section 3.10). Separate roles have been defined for investigators that have been responsible for various aspects of the development of a data set. These roles reflect that EMAP monitoring has been completed using multiple teams of contractor and federal staff and that often no one investigator is responsible for all activities related to a data set. The fields may be repeated as many times as is necessary to represent multiple investigators. This section of the catalog references investigator names only; additional information about each named individual is provided in the Personnel Information section of the catalog (Section 3.13)

Table 3-2. Catalog components identifying investigators

Field	Field Description
Princ Invest:	Principal Investigator
Sam Coll Invest:	Sample Collection Investigator
Sam Proc Invest:	Sample Processing Investigator
Data Anal Invest:	Data Analysis Investigator
Add Invest:	Additional Investigator

Discussion: These fields identify the investigators who had some active role in producing the data set. Those actively concerned with the generation and use of the data set are intended to be listed here; individuals responsible for the administration and management of the program are more appropriately referenced in the data set contact fields provided in the directory.

Although highly desirable, it is not necessary to provide information for principal investigators to complete a data set catalog entry. If principal investigator information is missing, the name of the contact person referenced in the directory will be given. If no contact person is identified, the name of the data center originating the data set will be given.

3.3 DATA SET ABSTRACT

A brief summary of the data set, like the abstract of a paper, allows a potential user to browse without learning all of the details. The General Keyword field can be repeated as many times as necessary.

Table 3-3. Catalog components describing abstract information

Field	Field Description
Abstract:	Abstract of the Data Set
Gen Keyword:	Keywords for the Data Set

Discussion: These fields are used to provide an overview of the data set. The Abstract should be a paragraph or two and summarize the main points expanded in the following sections. Any appropriate keywords should be listed in the Keyword field. A menu of suggested keywords is provided through the data set directory; however, catalog authors are not limited to these keywords.

3.4 OBJECTIVES AND INTRODUCTION

Information in this section provides background and justification for each data set within the context of the study from which it originated (Table 3-4). This material can usually be abstracted from research proposals or project plans. The Program Objective field will describe overall EMAP program objectives using standard text supplied by the Information Management Staff, and need not be completed by the catalog author.

Table 3-4. Catalog components describing objectives and introduction information

Field	Field Description
Prog Objectiv:	Program Objective
Data Objectiv:	Data Set Objective
Data Backgrd:	Data Set Background Information
Parameter Sum:	Summary of data set parameters

Discussion: These fields are roughly the equivalent of an introductory section in a scientific paper. They consist of a paragraph or two of textual material that gives an overview of the program and data set. These fields must be provided to complete a data catalog entry.

The Data Objective field is used to describe the specific goals that were to be met by the collection of this data set. The Data Set Background Information, on the other hand, sets this data collection activity in a broader context of scientific motivation and interactions with other data available or to be collected. The last field provides a summary (not an exhaustive list) of the parameters, variables, and measured quantities reported in the data set.

3.5 METHODS

The methods employed to acquire the data set are summarized in this section of the catalog. Documents and reports containing more extensive descriptions of methods, including standard operating procedures (e.g., Conkling and Byers, 1993; Macauley and Summers 1991) are more appropriately stored separately from the catalog, but are referenced in the data set catalog.

Descriptions of methods are organized in two subsections: data acquisition (including sampling; Table 3-5) and data preparation (including sample processing; Table 3-6).

3.5.1 Data Acquisition

Table 3-5. Catalog components describing data acquisition methods

Field	Field Description
Samp Objectiv:	Sampling Objective
Sampl Method:	Sample Collection Methods Summary
Beg Sampl Date:	Beginning Sampling Date
End Sampl Date:	Ending Sampling Date
Platform:	Sampling Platform
Sample Equip:	Sampling Equipment
Equip Manufac:	Manufacturer of Sampling Equipment
Key Variables:	Key Variables
Sam Meth Cal:	Sampling Method Calibration
Sam Qual Con:	Sample Collection Quality Control
Sam Meth Ref:	Sample Collection Method Reference
Sam Meth Dev:	Sample Collection Method Deviations

Discussion: These text fields describe why, how, and when sampling data were acquired. In general, they would consist of a sentence or two to possibly a paragraph each, forming the equivalent of a sampling methods description in a scientific paper.

For the most part, the information required is straightforward. The methods summary need not be exhaustive, as long as an adequate bibliographic reference to the complete description of the sample collection methods is included. (It is anticipated that the data set catalog will be linked to extensive text files containing complete descriptions of methods and standard operating procedures.) The most critical field may be the last, which is used to document any known deviations from the methods referenced. Knowledge of such deviations may be necessary to assess the overall quality of the data or to evaluate the significance of any long-term changes identified from the integration of separate data sets collected in multiple years.

3.5.2 Data Preparation and Sample Processing

Table 3-6. Catalog components describing data preparation and sample processing methods

Field	Field Description
Proc Objectiv:	Data Preparation Objective
Data Proc Sum:	Data Processing Methods Summary
Sampl Proc Calib:	Sampling Processing Method Calibration
Proc Qual Con:	Sample Processing Quality Control
Samp Proc Ref:	Sample Processing Method Reference
Sampl Proc Dev:	Sample Processing Method Deviations

Discussion: These text fields describe why and how samples were processed. In general, they would consist of a sentence or two to possibly a paragraph each, forming the equivalent of a sample processing description in a scientific paper.

For the most part, the information required is straightforward. The processing summary need not be exhaustive, as long as an adequate bibliographic reference to the complete description of the sample processing procedures is included. (It is anticipated that the data set catalog will be linked to extensive text files containing complete descriptions of methods and standard operating procedures.) The most critical field may be the last, which is used to document any known deviations from the standard operating procedures referenced. Knowledge of such deviations may be necessary to assess the overall quality of the data or to evaluate the significance of any long-term changes identified from the integration of separate data sets collected in multiple years.

3.6 DATA MANIPULATIONS

This section of the data set catalog provides documentation for any manipulations of the data subsequent to data acquisition and data preparation (Table 3-7). Documentation of these data manipulations should be sufficiently detailed so that future users of the data will fully understand what was done. The fields used in this section may be repeated as many times as is necessary to document the data manipulations used to produce the final data set.

Table 3-7. Catalog components describing data manipulations

Field	Field Descriptions
Deriv Value Name:	Name of New or Modified Value
Data Manip Desc:	Data Manipulation Description
Data Manip Exmpl:	Data Manipulation Examples
Man Code File:	Data Manipulation Computer Code File
Man Code Lang:	Data Manipulation Computer Code Language
Manip Code:	Data Manipulation Computer Code

Discussion: These fields describe any algorithms that have been used to derive new values or quantities from existing data. Such data manipulations may include, for example, conversions to different units, transformation of continuous data values to discrete values, or derivation of unmeasured quantities that are uniquely determined by known data. The data manipulation description and example fields must be completed and include complete descriptions of algorithms or reference information (e.g., look up tables) applied.

It may be necessary to use a "pseudo-code" approach to define the algorithm in a stepwise fashion in ASCII text. It is optional, but encouraged to include relevant program segments of the actual computer code used and to submit the program file for archiving with the data.

3.7 DATA DESCRIPTION

The data description sections give details of each value reported in the data set, sufficient for a user to read the data set file and ingest it into application software. The description is provided in three parts: 1) A description of the quantities in the data set (Table 3-8), 2) An example data record (Table 3-9), and 3) A cross reference to related data sets (Table 3-10). The data description contents and accuracy are primarily a function of the investigator submitting the data set. However, the description provides the basis for organizing the data itself in the EMAP relational data base management system (RDBMS), and will be entered into the RDBMS data dictionary and used heavily by the information management staff. This usage will generate feedback to the investigator and to this catalog document through efforts to standardize like parameter names and definitions across data sets, checks of maximum and minimum values, reports generating data record examples, and identification of related data sets.

3.7.1 Description of Parameters

Parameters in the data set are listed and described in this section of the data catalog. This information is abstracted from the data dictionary for the EMAP monitoring data base and may be presented in tabular form. The following information should be provided for each quantity (column) in the data set:

Table 3-8. Catalog components describing data set parameters

Field	Field Descriptions
Param Name:	Parameter Name
SAS Param Name:	SAS Parameter Name
Param Descrip:	Parameter label or description
Units:	Units of measurement
Data Type:	Parameter data type
Precision:	Precision to which values are reported
Accuracy:	Accuracy of the data values
Act Min Value:	Minimum Value in Data Set
Act Max Value:	Maximum Value in Data Set

Discussion: These fields describe the basic characteristics (name, type, accuracy, range) of values that are reported in the data set. The minimum value and maximum value fields are optional for non-numeric values. The shortened (8 character) SAS name should be provided if the data set or the parameters in the data base originated from a SAS data set. The parameter description is a one line complete and precise scientific name of the quantity, not an extensive definition or algorithm.

3.7.2 Data Record Example

A display of a limited number of records or observations in the data set assists the data user in understanding the structure and composition of the data set. This section of the data set catalog presents several example records from the data set. The records to be displayed can be provided by the investigator or can be generated from the EMAP relational data base after the data set is loaded into the data base. If the data set is complex or there are specific examples that should be shown, the investigator should elect to provide the material at the time the catalog document is written. The Example Data Values line is repeated once for each data set record included.

Table 3-9. Catalog components describing data record

Field	Field Description
Header Line:	Column Names for Example Records
Exmpl Data Values:	Example Data Records

Discussion: The example data values from a data set record are recorded in this field. They should be spaced so as to line up with the column names in the header line, when displayed in a fixed (non proportional) font.

Special arrangements should be made if this field is to be used in the description of binary data sets (for example, SAS files, GIS coverages, or a satellite images). The field serves two purposes: to give the user a preview of the scientific data and to give the programmer who must access the data formatting and verification information. For the first purpose, if numbers are relevant, they can be extracted, and a "pseudo record" constructed. Similarly, a thumbnail version of a displayed GIS or image data set could be attached to the catalog document. Actual byte values can be included to meet the second purpose, if necessary. For example, a TIFF image header and the corresponding values in the data file could be listed in the two fields.

3.7.3 Related Data Sets

The names and identification codes of data sets containing similar or related data are referenced in this section of the data catalog (Table 3-10). The references should be entered exactly to make possible relational links (via the RDBMS) to the documentation for those data sets. These fields can be repeated to accommodate multiple data set references.

Table 3-10. Catalog components describing data record

Field	Field Description
Related DS Name:	Related Data Set Name
Task Group:	Task Group Code
Related DS ID:	Related Data Set Identification Code
Version:	Version number

Discussion: The name and identification code of a data set containing similar or related data is recorded in this field.

3.8 GEOGRAPHIC AND SPATIAL INFORMATION

Information about the spatial coverage of the data set, particularly information that is specific to spatial data sets, is provided in this section of the data set catalog (Table 3-11). In the geographic coverage section, specific spatial data organization and reference information will be supplied if applicable to the data set. (The spatial data documentation standards recommended by the Federal Geographic Data Committee (FGDC 1994) have been considered in designing this set of fields - see the Appendix for additional details.)

Table 3-11. Catalog components describing geographic and spatial information

Field	Field Description
Min Longitude:	Minimum Longitude
Max Longitude:	Maximum Longitude
Max Latitude:	Maximum Latitude
Min Latitude:	Minimum Latitude
Geo Keyword:	Name of the area or region
Spatial Ref Meth:	Direct Spatial Reference Method
Horiz Coord Sys:	Horizontal Coordinate System Used
Horiz Resolution:	Resolution of Horizontal Coordinates
Horiz Coord Units:	Units for Horizontal Coordinates
Vertical Coord Sys:	Vertical Coordinate System
Vertical Resolution:	Resolution of Vertical Coordinates
Vertical Coord Units:	Units for Vertical Coordinates

Discussion: These fields provide basic information about the spatial extent and organization of the data set. The maximum and minimum longitude and latitude give the bounding coordinates (West, East, North, South) marking the limits of coverage of a data set. Latitude and longitude values are expressed in decimal degrees. The geographic keywords or names should uniquely identify the spatial extent of the data set using a term descriptive of the boundaries (e.g. New York State; EPA Region IX) or geographic features (e.g. Chesapeake Bay; Virginian Province). This field may be repeated to indicate multiple overlapping region names. The spatial reference method informs the user whether the data set organization is point, vector, or raster.

3.9 QUALITY CONTROL/QUALITY ASSURANCE

Quality control and quality assurance information is used to understand the limits of the data. Specific data collected to assess data quality may be included in this section of the data set catalog (Table 3-12), or may be included in a separate data set that is referenced in the data set catalog.

Table 3-12. Catalog components describing quality control/quality assurance information

Field	Field Description
Meas Qual Obj:	Measurement Quality Objectives
QA/QC Meth:	Quality Assurance/Control Methods
Act Meas Quality:	Actual Measurement Quality
Sources of Error:	Sources of Error
Known Data Prob:	Known Problems with the Data
Conf Level Stmnt:	Confidence Level/Accuracy Judgement
Allow Min Value:	Allowable Minimum Values
Allow Max Value:	Allowable Maximum Values
QA Ref Data:	QA Reference Data

Discussion: In these fields the investigator should provide brief discussions of the quality issues associated with the data set. The first three fields encode the formal quality objectives, methods, and results obtained. The next set of fields should contain more descriptive information such as field notes on sources of error, problems encountered by the investigator or others in using the data set for analysis, and a subjective evaluation of the original investigator's confidence in the data set. The allowable minimum and maximum values are useful QA check and search query information. The last field may include actual reference data or the name of a file or document that has the appropriate reference data.

3.10 DATA ACCESS

This section is designed to provide users with information on how to access data (Table 3-13).

Table 3-13. Catalog components describing data access information

Field	Field Description
Data Access:	Data Access Procedures
Data Access Restrict:	Data Access Restrictions
Data Access Contact:	Data Access Contact Person
Data Set Format:	Data Set Format
FTP Infor:	Information Concerning Anonymous FTP
Gopher Infor:	Information Concerning Gopher
WWW Infor:	Information Concerning World Wide Web
EMAP CD-ROM:	EMAP CD-ROM Containing the Data set

Discussion: The procedures discussion should provide general information about the different ways users can access data including telephone contact, dial-in lines, and Internet. Access to data via Internet most likely will utilize standard file transfer protocols (FTP), as well as the Internet discovery and retrieval tools Gopher and World Wide Web (WWW) provided through the Agency's public access server (Strebel and Frithsen 1995). Additional reference may be made to the EMAP Internet server for draft documents and data; however, access to this server is restricted and these restrictions will need to be documented. Contact information is entered in a descriptive way in these fields, but the content should be essentially the same as required for the investigator name and address section.

3.11 REFERENCES

This section of the data set catalog provides a list of any published documentation relevant to the data collected. Documentation may include manufacturer's instruction manuals, government technical manuals, user's guides, etc. Also referenced should be any technical reports and scientific publications concerning the methods, instruments, or data described in this document. Publications by the principal investigator or investigating group that would help a reader understand or analyze the data are particularly important. The format of the bibliographic references is taken from the EMAP bibliographic data base specification (Lear, pers. comm.). The section is broken into two subsections to accommodate references that will be maintained in the relational tables for the EMAP bibliographic data object (detailed reference format) and those background and general references which are primarily of use to the reader and will not be tracked (brief reference format).

3.11.1 EMAP References

References given in this section are those published or used as part of EMAP and which should be tracked in the data table for the EMAP bibliographic data object (Table 3-14).

Table 3-14. Catalog components describing EMAP reference information

Field	Field Description
Ref Type:	Reference Type
Ref Author:	Reference Author
Ref Authors Affil:	Reference Author's Affiliation
Ref Title:	Title of Reference
Volume Title:	Journal or Volume Title
Volume Editor:	Journal or Volume Editor
Page Ref:	Page and Volume Reference
Date of Ref:	Date the Reference was Published
Place of Pub:	Location of Publishing Organization
Publisher:	Name of Publishing Organization
Ref Other ID:	Reference Report Number or Other ID
Procite Rec Num:	Procite Record Number for the Reference

Discussion: Standard bibliographic reference information is provided in these fields. The type of documentation referenced may include journal articles, workshop proceedings, books, reports, films, video tapes, audio tapes, CD-ROMs,etc. All authors are listed using paired repeats of the fields Reference Author and Reference Authors Affiliation. If the reference is an article in a journal, a proceedings volume, or other compendium, the title and editor of the full work are given in the "Volume" fields. For special reports, technical memoranda, or other material that carries an internal organizational identification code or report number, it should be recorded in the Reference Other Identification field. The Procite Record Number is optional. It is included in the catalog specifications because the EMAP bibliography is currently maintained in a Procite data base. Eventually, all EMAP bibliographic information should be stored and maintained in the EMAP Oracle data base, eliminating the effort needed to maintain separate data bases.

3.11.2 Background References

The references listed in this section are materials in the open literature which may support the collection, use, and interpretation of the data set, but are not directly related to or products of EMAP. As such they need not be tracked in the EMAP bibliographic data object, but should be captured as part of the documentation. A single text field is provided, but it should contain all of the relevant information that would otherwise have been listed in the previous section (Table 3-15).

Table 3-15. Catalog components describing EMAP background references

Field	Field Description
Supt Ref:	Supporting Reference

Discussion: This field contains the reference information as described above for a single supporting reference relevant to the data set. It may be repeated as many times as required.

3.12 GLOSSARY AND TABLE OF ACRONYMS

The detailed documentation for each data set are likely to contain terms and acronyms that are unfamiliar to some users of the data. Catalog authors should provide definitions for those terms that have specific meaning within EMAP or within a particular discipline (Table 3-16). All acronyms used should be recorded in these fields (which are repeated in pairs) to provide convenient references for detailed documentation reports and on-line documentation searches.

Table 3-16. Catalog components describing glossary terms and acronyms used in the catalog entry

Field	Field Description
Acronym:	Glossary term or acronym used in the Detailed Documentation
Acronym Def:	Definition of glossary term in acronym

Discussion: These fields provide definitions for terms and acronyms in the catalog that may be unfamiliar to some users of the data. The EMAP Master Glossary (EMAP 1993) should be consulted for suggested definitions for terms that have specific meaning within EMAP.

3.13 PERSONNEL INFORMATION

This section of the data set catalog contains information identifying the individuals associated with the data set and named in the data set catalog (Table 3-17). This includes catalog authors (Section 3.1), investigators (Section 3.2), and data set contacts (Section 3.10). Whereas names only were provided in previous sections, this section provides address, telephone and email information for individuals named within the catalog. The fields used to provide this information are the same fields used in the data set directory and the contacts data base object. Individuals in the contacts data base object are linked to the catalog through their defined roles and the data set identification information.

Table 3-17. Catalog components describing personnel information.

Title:	Formal Title
Lst Name:	Last Name
Frst Name:	First Name
Mid Init:	Middle Initial
Role	Role
Address1:	Line 1 of address
Address2:	Line 2 of address
Address3:	Line 3 of address
Address4:	Line 4 of address
City:	City
State:	State
Zip:	Zip Code
Country:	Country
Voice Phone:	Voice Phone Number
Fax Phone:	Fax Phone Number
Email Addr:	Email Address
Email Netwk:	Email Network
Add EM Inf.	Additional Email Information

4.0 EXAMPLES

The purpose of this chapter is to illustrate how typical data set information may be associated with the fields of the catalog. The format in this chapter would be convenient for writing the entry, updating it , submitting it to the IMS staff, and loading it into the IMS documentation data base. It is not intended to be the output format seen by data users and others who consult the data set documentation. The output formats are discussed in Chapter 5.

The example entries are drawn from information obtained from the EMAP Estuaries Resource Group. Example catalog documentation is provided for a biological data set (1990 Virginian Province Benthic Data Set) and a chemical data set (1990 Virginian Province Sediment Metal Chemistry Data Set). The catalog documentation examples are provided as illustrations only and for this draft have not been reviewed by resource group scientists or information management specialists from the EMAP Estuaries Resource Group. Although attempts were made to complete accurate catalog entries, some information was not available for this draft and hypothetical information was provided to complete the example. Readers are cautioned not to use these examples as working documentation for the data sets described. Note: The examples were incompletely developed for the 30 April 1995 draft of this report.

4.1 BIOLOGICAL DATA SET

This example describes a benthic data set collected in 1990. Although it is based on an actual example, some information is hypothetical to complete the example.

4.1.1 Data Set Identification

Title:	1990 Virginian Province Benthic Data for Species Abundance and Biomass
Cat Author:	Jeffrey B. Frithsen
Cat Rev Date:	1995-04-01
Data Set Name:	EMAP - Esturaries Program Level Database - 1990 Virginian Province Benthic Species Data Set Summarized by Station
Task Group:	01
Data Set Id:	1001
Version:	001
Req Acknowl:	These data were produced as part of the U.S. EPA's Environmental Monitoring and Assessment Program (EMAP). Although the data described in this article have been funded wholly or in part by the U.S. Environmental Protection Agency through its EMAP Estuaries Program, it has not been subjected to Agency

review, and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

4.1.2 Investigator Information

Princ Invest:	
Sam Coll Invest:	Charles J. Strobel
Sam Proc Invest:	Jeffrey B. Frithsen
Data Anal Invest:	A. Frederick Holland
Add Invest:	Steve Weisberg

4.1.3 Data Set Abstract

Abstract: This data set contains measurements for the abundance and biomass of bottom-dwelling (benthic) macroinvertebrates collected from sediments within the estuaries of the Virginian Province (Cape Code to Chesapeake Bay). All samples were collected during the summer using a young-modified Van Veen grab. Abundance measurements were made for individual taxa; biomass measurements were made for selected species and taxonomic groups.

Gen Keyword:	Benthic
Gen Keyword:	Macroinvertebrate
Gen Keyword:	Estuary

4.1.4 Objectives and Introduction

Prog Objectiv: The EPA's Environmental Monitoring and Assessment Program (EMAP) was designed to provide quantitative assessment of the national extent of environmental problems by measuring status and change in selected indicators of ecological condition.

Data Objectiv: The specific objective of the investigation described in this document was to collect information to characterize the bottom dwelling (benthic) animal assemblages in the estuaries of the EMAP-Estuaries Virginian Province.

Data Backgrd: Macrobenthic organisms play an important role in the estuarine conceptual model. As major secondary consumers in estuarine ecosystems, they represent an important linkage between primary producers and higher trophic levels for both planktonic and detritus- based webs. They are a particularly important food source for juvenile fish and crustaceans and also include many commercially and recreationally important species.

The benthic macroinvertebrate species composition and abundance indicator has been placed in the core group not only because of its importance, but also because of its responsiveness to the kinds of environmental stress gradients of interest to EMAP-E. Benthic assemblages are composed of diverse taxa with a variety of reproductive modes, geeding guilds, life history characteristics, and physiological tolerances to environmental conditions. As a result, benthic populations respond to changes in conditions, both natural and anthropogenic, in a variety of ways. Responses of some species (e.g., filter feeders, species with pelagic life stages) are indicative of water quality changes, while responses of others (e.g., organisms that burrow in or feed on sediments) are indicative of changes in sediment quality.

Parameter Sum: Benthic species composition, abundance, and biomass were estimated for each of three sediment grabs taken at each sampling station.

4.1.5 Methods

4.1.5.1 Data Acquisition

Samp Objectiv: To collect sediment samples suitable for the analysis of benthic assemblage characteristics. Three replicate sediment samples were to be collect as each EMAP-VP station for benthic species composition, abundance, and biomass.

Sampl Method: Sediment grabs used for benthic samples were randomly interspersed with the grabs used for sediment chemistry/toxicity samples.

Beg Sampl Date: 1990-06-20
End Sampl Date: 1990-09-22
Platform: Samples were collected from 8-m (24 ft) twin-engine Chesapeake style workboats.

Sample Equip: A 1/25 m^2 stainless stell Young-modified Van Veen grab sampler with a maximum penetration depth of 10 cm was used to collect sediment samples. The sampler was constructed entirely of stainless steel and had been Kynar (similar to Teflon) coated.

Equip Manufac: Young: Falmouth, MA

Key Variables: RPD Depth;
Sam Meth Cal: The sample gear required no calibration.

Sam Qual Con: Acceptable grabs penetrated the sediments at least 7 cm. Grabs containing no sediments partially filled grabs, or grabs with grossly slumped surfaces were unacceptable. Grabs completely filled to the top where the sediment was in direct contact with the hinged top were also unacceptable.

Sam Meth Ref: Strobel, C.J. 1990. Environmental Monitoring and Assessment Program - Near Coastal Component: 1990 Demonstration Project Field Operations Manual. Environmental Research Laboratory, Narragansett, RI. U.S. Environmental Protection Agency, Office of Research and Development.

Sam Meth Dev: None.

4.1.5.2 Data Preparation and Sample Processing

Proc Objectiv: To process sediment samples to characterize benthic assemblages in terms of species composition, abundance, and biomass.

Data Proc Sum: Benthic fauna identified included those commonly termed "macrofauna" by benthic ecologists. "Meiofaunal" groups were not identified or enumerated. These groups included: nematodes, ostracods, turbellarians, harpacticoid copepods and foraminifera. In addition to meiofauna, taxonomic groups having only planktonic forms were excluded from the identification process. Examples of these groups were copepods and cladocerans.

Benthic fauna were identified to the lowest practical taxonomic level. Macrobenthos were identified to species, except for the following groups: class Anthozoa (class), subclass Copepoda (order), phylum Nemertinea (phylum), subclass Ostracoda (subclass) and class Turbellaria (class). For samples collected in low salinity (less than 5 ppt) water, oligochaets were identified to species and chironomides to genus. Above 5 ppt salinity, oligochaetes were identified to class and chironomids were identified to family.

Sampl Proc Calib: No calibration required.

Proc Qual Con: Ten percent of all samples were resorted as a quality control check on each technician's efficiency.

Samp Proc Ref: Klemm, D.J., L.B. Lobring, J.W. Eichelberger, A. Alford-Stevens, B.B. Porter, R.F. Thomas, J.M. Lazorchak, G.B. Collins, and R.L. Graves. 1993. Environmental Monitoring and Assessment Program (EMAP) Laboratory Methods Manual: Estuaries. U.S. Environmental Protection Agency, Environmental Research Laboratory, Cincinnati, OH.

Frithsen, J.B. 1991. Technical scientific assistance to EMAP Near Coastal Group for the processing and integration of EMAP-NC Demonstration Project benthic data. 6 May 1991. Report completed for the U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Washington, DC. Report prepared by Versar, Inc., Columbia, MD.

Sampl Proc Dev:

4.1.6 Data Manipulations

Deriv Value Name: BSPECNUM
Data Manip Desc: BSPECNUM (Total number of individuals of each species code = abundances of a species code summed across 'n' grabs collected at a station, where n=1to 3.
Data Manip Exmpl:
Man Code File:
Man Code Lang:
Manip Code:

Deriv Value Name: BSPEC_MA
Data Manip Desc: BSPEC_MA (mean abundance of indivuals for each species) = Abundances of a species code summed over 'n' grabs and divided by 'n' grabs

Dat manip Exmpl:
Man Code File:
Man Code Lang:
Manip Code:

Deriv Value Name: BSPECSTD
Data Manip Desc: BSPECSTD (standard deviation of the mean abundance) = Standard deviation of the mean abundance of each species code
Dat Manip Exmpl:
Man Code File:
Man Code Lang;
Manip Code:

4.1.7 Data Description

4.1.7.1 Description of Parameters

Param Name:
SAS Param Name: STA_NAME
Param Descrip: Station identification code
Units:
Data Type: Character
Precision:
Accuracy:
Act Min Value:
Act Max Value:

Param Name:
SAS Param Name: VST_DATE
Param Descrip: Date sampling started at this station
Units:
Data Type: Date (YYMMDD)
Precision:
Accuracy:
Act Min Value: 90-06-20
Act Max Value: 90-09-22

Param Name:
SAS Param Name: SCINAME
Param Descrip: Scientific name of the species or taxonomic group
Units:
Data Type: Character
Precision:

Accuracy:
Act Min Value:
Act Max Value:

Param Name:
SAS Param Name: BSPECNUM
Param Descrip: Number of individuals for a specific value of SCINAME
Units: Number per 440 cm^2 grab
Data Type: Numeric
Precision: 1
Accuracy: +- 10 %
Act Min Value: 0
Act Max Value:

Param Name:
SAS Param Name: BSPEC_MA
Param Descrip: Mean number of individuals for a specific value of SCINAME for all grabs collected at a station
Units: Mean number per 440 cm^2 grab
Data Type: Numeric
Precision: 1
Accuracy:
Act Min Value: 0
Act Max Value:

Param Name:
SAS Param Name: BSPECSTD
Param Descrip: Standand deviation of the mean number of individuals for a specific value of SCINAME for all grabs collected at a station
Units:
Data Type: Numeric
Precision:
Accuracy:
Act Min Value:
Act Max Value:

4.1.7.2 Data Record Example

Header Line:
Exmpl Data Values:

4.1.7.3 Related Data Sets

Related DS Name:	
Related DS ID:	

4.1.8 Geographic and Spatial Information

Min Longitude:	77.17048
Max Longitude:	70.04186
Max Latitude:	41.38330
Min Latitude:	36.49546
Geo Keyword:	Virginian Province
Geo Keyword:	EPA Region I
Geo Keyword:	EPA Region II
Geo Keyword:	EPA Region III
Spatial Ref Meth:	Point
Horiz Coord Sys:	Geographic
Horiz Resolution:	
Horiz Coord Units:	
Vertical Coord Sys:	
Vertical Resolution:	
Vertical Coord Units:	

4.1.9 Quality Control/Quality Assurance

Meas Qual Obj: Measurement quality objectives were outlined in the Quality Assurance Project Plan (Valente et al. 1991). The objective of laboratory processing was to achieve a maximum error for sorting, counting, and taxonomic identifications of 10% of the individuals in the sample.

QA/QC Meth: Quality control for processing grab samples involved both sorting and counting check systems. A check on the efficiency of the sorting process was required to document the accuracy of the organism extraction process. In addition to sorting QA, it was necessary to perform checks on the accuracy of sample counting. This was done in conjunction with taxonomic identification and used the same criteria presented for taxonomic identification quality control.

Act Meas Quality: Actual measurement quality as indicated by quality assurance audits indicated the error for sorting as 3.1% of the number of specimens in the sample. The error for species identification and enumerations was 1.4%.

Sources of Error: The methods used to process benthic samples require that a small number of representative specimens of each species be set aside in a taxonomic reference collection. Therefore, the total biomass is underestimated for those samples from which reference specimens were taken.

Total macrofaunal biomass was also potentially underestimated for samples from tidal freshwater and oligohaline salinity regions where the number of chironomids or the number of oligochaetes was less than 20. In those instances where the number of oligochaetes or chironomids was less than 20, all specimens were mounted for identification and no biomass measurements were made.

Known Data Prob: Known data problems were not judged to be significant but were documented in the report describing the laboratory processing of benthic samples (Frithsen 1991).

Conf Level Stmnt:
Allow Min Value:
Allow Max Value:
QA Ref Data:

4.1.10 Data Access

Data Access:
Data Access Restrict:
Data Access Contact:
Data Set Format:
FTP Infor:
Gopher Infor:
WWW Infor:
EMAP CD-ROM:

4.1.11 References

4.1.11.1 EMAP References

Ref Type:	Manual
Ref Author:	Klemm, D. J.
Ref Authors Affil:	
Ref Author:	Lobring, L. B.
Ref Authors Affil:	
Ref Author:	Eichelberger, J. W.
Ref Authors Affil:	
Ref Author:	Potter, B. B.
Ref Authors Affil:	
Ref Author:	Thomas, R. F.
Ref Authors Affil:	
Ref Author:	Lazorchak, J. M.
Ref Authors Affil:	
Ref Author:	Collins, G. B.
Ref Authors Affil:	
Ref Author:	Graves, R. L.
Ref Authors Affil:	
Ref Title:	Environmental Monitoring and Assessment Program (EMAP) Laboratory Methods Manual: Estuaries.
Volume Title:	
Volume Editor:	
Page Ref:	
Date of Ref:	September, 1991
Place of Pub:	Cincinnati, Ohio
Publisher:	U.S. EPA Office of Research and Development, Environmental Monitoring Systems Laboratory
Ref Other ID:	EPA/600/4-91/xxx
Procite Rec Num:	

4.1.11.2 Background References

Supt Ref:	

4.1.12 Glossary and Table of Acronyms

Acronym:	ppt
Acronym Def:	parts per thousand

4.1.13 Personnel

Title:	Dr.
Lst Name:	Frithsen
Frst Name:	Jeffrey
Middle Init:	B.
Role:	Catalog Author
Address1:	Versar, Inc.
Address2:	9200 Rumsey Road
Address3:	
Address4:	
City:	Columbia
State:	MD
Zip:	21045
Country:	USA
Voice Phone:	410-740-6112
FAX Phone:	410-964-5156
Email Address:	Frithsenjef@Versar.Com
Email Network:	Internet
Add EM Info:	

4.2 CHEMICAL DATA SET

This example was not included in the April 30, 1995 draft of this report.

5.0 CATALOG OUTPUT TO USERS

The catalog fields are designed to reflect the elements of a scientific paper. The reader should be given a document-like text format that is easy to read and understand. There should also be a capability to search a document by section. This chapter gives guidance on constructing those output formats, but does not explore implementation details.

5.1 THE METADATA PUBLICATION

The catalog sections are to be presented as the sections of a paper. Each section header should match the section number and title (to permit searching the file electronically). Within each section, the fields names are to be deleted and the field contents are to be aggregated into complete sentences and paragraphs that can be easily read. Some sections will have more descriptive text than others. For example, the first section contains brief identification information and might be formatted (compare 4.1.1 above):

Section 1: Data Set Identification

1990 Virginian Province Benthic Data for Species Abundance and Biomass
This document was written by Jeffrey B. Frithsen
This document was last revised on April 1, 1995
This document describes the EMAP-E Estuaries-1990 Virginian Province Benthic Species Data Set Summaried by Station
The data set identification number is 1001, Version 001.
The following acknowledgment is requested by the investigators of anyone who uses and publishes on these data:
These data were produced as part of the U.S. EPA's Environmental Monitoring and Assessment Program (EMAP). Although the data described in this article have been funded wholly or in part by the U. S. Environmental Protection Agency through its EMAP Estuaries Program, it has not been subjected to Agency review, and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

On the other hand, a section with a significant amount of text already provided in descriptive format would simply be broken into paragraphs. For example, consider this section about quality control (compare 4.2.9 above):

Section 9: Quality Control/Quality Assurance

Measurement Quality Objectives (MQOs) for the 1990 Virginian Province sediment chemistry analyses were defined in the 1990 Demonstration Project Quality Assurance Project Plan for EMAP-Near Coastal (Valente, et al., 1990). This plan required each laboratory to analyze the *following quality control (QC) samples along with every batch or "set" of field chemistry samples: laboratory reagent blank, calibration check standards, laboratory fortified sample matrix, laboratory duplicate, and Laboratory Control Material (LCM). Results for these QC samples had to fall within certain pre-established control limits for the analysis of a batch of samples to be considered acceptable.*

Results of QC sample analyses are stored in the EMAP-Estuaries data base and are available upon request. For the analysis of major and trace elements by ICP-AES and GFAA, the laboratory generally met the pre-established acceptability criteria (control limits) for the QC samples. For the 1990 mercury analyses, the average percent recovery in the reference material fell just outside the accuracy control limit range of 85% to 115%, suggesting that mercury may have been slightly under-recovered in some sample batches. All QC results for the analysis of total organic carbon in the 1990 sediment samples fell within required control limits.

A major deficiency in the 1990 organics data set is related to the laboratory's failure to achieve the target detection limits originally specified in the QA plan. These target detection limits were 10ng/g (dry weight) for each PAH compound and 0.5 ng/g for each PCB congener and pesticide. In general, the detection limits achieved by the laboratory ranged from 1.5 to 30 times higher than the target value for PAH compounds and up to 15 times higher than the target value for PCB congeners and pesticides. In addition, the detection limits varied widely because the laboratory analyzed a different amount (i.e., dry weight) of sediment from each sample. As a result, the analytes of interest were not detected in a large number of samples, and the "calculated detection limit (i.e., the theoretical concentration of each analyte necessary for detection) differed significantly from sample to sample.

Data users are cautioned that there are several major deficiencies in the 1990 sediment organics data set that might limit or preclude the use of these data. These deficiencies were the result of numerous methodological and QA/QC problems experienced by the laboratory responsible for the analysis. Data users are cautioned that there are deficiencies in the 1990 sediment data set for butyltin compounds which might limit or preclude the use of these data. The laboratory's failure to detect the butyltin compounds of interest (TBT, DBT, MBT) in the majority of samples analyzed suggests a potential deficiency from the method detection limits for the individual analytes. The method detection limit (MDL) established by the laboratory was 4 ng/g dry weight for both TBT and DBT and 10 ng/g dry weight for MBT. Assuming these MDLs are valid, it is probable that contamination by butyltin compounds may be more widespread than indicated by these data.

If the target detection limits had been achieved and consistent sample sizes had been used, the organic analytes of interest probably would have been detected and quantified in most of the 1990 Virginian Province samples. In reality, analytes of interest present in the samples at low concentrations were not detected and therefore not reported. This limits the comparability of the

1990 organics data with other data sets for which lower detection limits were achieved and limits data users' ability to make quantiative evaluations of sediment contamination for these organic compounds in the Virginian Province. The 1990 mercury and TOC results were deemed acceptable for use without qualification.

Actual report formatting instructions will depend on how the documentation fields suggested in this document are implemented in a data base. This is an implementation issue which cannot be addressed at this time.

5.2 SELECTION OF METADATA COMPONENTS

A full document may be difficult to view and search electronically. For the EMAP IMS online system, a document should be viewable using a menu that allows the reader to select specific components or sections of the document. It is suggested that document sections formatted as described in 5.1 above be stored so that they may be accessed by search tools (such as gopher) or hypertext links (as in html) organized by the primary document sections. That is, the user should be given the document title and the following list of options from which to choose:

The components of the data set catalog are organized into 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. Glossary and Table of Acronyms
13. Personnel Information.

A metadata access system will be required to present the user with lists of metadata documents from which to choose. This may be based on the EMAP data set directory or organized hierarchically by resource group and data type in a manner parallel to the data directories on the public access server. The metadata access system is a separate design topic and is not discussed further in this document.

6.0 LITERATURE CITED

Conkling, B.L. and G.E. Byers (eds.). 1993. Forest Health Monitoring Field Methods Guide, Revised July, 1993. Internal Report. EPA/600/X-92/073. U.S. Environmental Protection Agency, Las Vegas, NV.

EMAP. 1993. Environmental Monitoring and Assessment Program: Master Glossary. EPA/620/R-93/013. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Research Triangle Park, NC.

EOSDIS 1994. Interim Release2 ECS Core Metadata Baseline. 194-00269TPW. EOSDIS Core System Project. Hughes Applied Information Systems, Inc., Landover. MD.

FGDC 1994. Content Standards for Digital Geospatial Metadata. June 8, 1994. Federal Geographic Data Committee. Washington, DC.

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, Office of Research and Development, Environmental Monitoring and Assessment Program, Washington, DC. Report prepared by Versar, Inc., Columbia, MD.

Justice, C.O., G.B. Bailey, M.E. Maiden, S.I. Rasool, D.E. Strebel and J.D. Tarpley. 1995. Recent data and information system initiatives for remotely sensed measurments of the land surface. Remote Sensing of the Environment 51:235-244.

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

Lear, J.S. and C.B. Chapman, eds. 1994. Environmental Monitoring and Assessment Program (EMAP Cumulative Bibliography. EPA/620/R-94/024. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Research Triangle Park, NC.

Lear, J.S. 1994. Personal communicaton. Components of the EMAP bibliographic data base. ManTech Environmental Technology, Inc., Research Triangle Park, NC.

Macauley, J.M. and J.K. Summers. 1991. Near Coastal Louisianian 1991 Province Demonstration Project - Field Operations Manual. Environmental Research Laboratory, Gulf Breeze, FL. U.S. Environmental Protection Agency, Office of Research and Development.

Meeson, B.W., D.E. Strebel and E.D. Paylor. 1993. Earth science information systems: A perspective from the Pilot Land Data System. In, A. Zygielbaum (ed) Earth and Space Science Information Systems, AMerican Institute of Physics Conference Proceedings 283, AIP Press, NY, PP. 216-226.

Michener, W.K., A.B. Miller, and R. Nottrott. 1990. Long-Term Ecological Research Network Core Data Set Catalog. Belle W. Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC.

NRC. 1991. Solving the global change puzzle: A U.S. Strategy for Managing Data and Information. A report by the Committee on Geophysical Data, Commission on Geosciences, Environment, and Resources, National Research Council. National Academy Press, Washington, DC.

Palmer, J.D. and A. Fields. 1994. Report on the Environmental Monitoring and Assessment Program, Surface Waters JAD Session, Surface Waters Facilities, Corvallis, OR, June 6-9, 1994. June 23, 1994. Report completed for the U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Washington, DC. Report prepared by George Mason University, Fairfax, VA.

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., B.W. Meeson, and A.K. Nelson. 1994a. Scientific information systems: A conceptual framework. *In:* Environmental Information Management and Analysis: Ecosystem to Global Scales. W.K. Michener, J.W. Brunt, and S.G. Stafford, eds. Taylor and Francis, Bristol, PA.

Strebel, D.E., D.R. Landis, K.F. Huemmrich and B.W. Meeson. 1994b. Collected Data of The First ISLSCP Field Experiment, Volume 1: Surface Observations and Non-Image Data Sets. Published on CD-ROM by NASA.

USEPA. 1993a. 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, Environmental Monitoring and Assessment Program (EMAP), Washington, DC.

USEPA. 1993b. System Design Specifications for the Proof of Concept (POC). February 26, 1993. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Assessment Program (EMAP), Washington, DC.

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.

APPENDIX
DETAILED SPECIFICATIONS FOR DATA SET CATALOG CONTENTS

This appendix is intended to be a field-by-field technical reference for those concerned with finalizing, formatting, and submitting catalog documents to the EMAP Information Management System. Note that, although all information is entered and stored in individual fields for entry into the data base management system, the information is designed to be provided to potential users in a document-like report format that is easy to read and understand (see Chapter 5).

A.1 DATA SET IDENTIFICATION

This section of the catalog contains information used to identify the data set. Most of this information is taken directly from the data set directory.

Title: Title

Description: This field contains the title of the data set catalog entry.

Recommendation: The title of the data set catalog entry, in most cases, will be similar to the title of the data set. The field should be a variable-length text field. Typically the title is no more than 160 characters.

Cat Author: Catalog Author

Description: The name or names of the author(s) of the catalog document are provided using this field. One name, last name first, should be entered into each instance of the field. The field may be repeated as many times as necessary. Note that those making significant revisions to an existing document should add their name to the author list as well as update the Catalog revision date.

Recommendation: This should be a variable length character field.

Cat Rev Date: Catalog revision date

Description: The scientific documentation of a data set will change over time to reflect the results of new analyses and to incorporate information concerning additional data uses, limitations, and publications. The catalog revision date field is used to record when the catalog entry was last revised.

Recommendation: This field is mandatory. Dates should be given in the format YYYY-MM-DD, where DD is the two-digits for the date, MM is the two digits signifying the month, and YYYY is the four-digit year. Leading zeroes are used in the entry if needed.

Data Set Name: Data set name

Description: This field contains a descriptive name for a data set. This field is also included in the directory.

Recommendation: The format of this field is specified in the guidance provided for building directory entries.

Task Group: Task Group

Description: Name of the EMAP task group from which the data set originates.

Recommendation: This field is mandatory. EMAP task groups are referenced using a unique two-digit code. Valid codes for each task group are given in Table A-1.

Table A-1. Valid entries for EMAP Task Groups

01	Estuaries
02	Forests
03	Surface Waters
04	Agricultural Lands
05	Rangelands
06	Great Lakes
07	Landscape Ecology
08	Wetlands
09	Assessment
10	Design and Statistics
11	Information Management
12	Indicators
13	Integration
14	Landscape Characterization
15	Logistics and Methods
16	Stressors
17	EMAP Center

Data Set Id: Data set identification code

Description: Number assigned by information management personnel within an EMAP task group to identify a data set. This field is mandatory. The data set identification is a positive whole number. Each task group will ensure that no data set from the same task group is assigned the same number. The Data Set ID and the Task Group will be concatenated to determine a unique identifier for each data set; therefore, different task groups having the same data set identification number will not present a problem.

Version: Version number for a data set

Recommendation: This field is mandatory. The version number should be a positive whole number.

Comment: Any change in a data set results in the creation of a new data set or an updated version of an old data set. All changes need to be documented and that documentation included as part of the data set catalog.

Req Acknowl: Requested acknowledgement

Description: Because the data set may be referenced in scientific publications, a suitable acknowledgement for its use should be suggested. Data sets collected by EMAP staff under programmatic auspices will have a general EMAP acknowledgement statement. Where a data set is the product of a specific principal investigator's work, and should be so referenced, the principal investigator should provide appropriate acknowledgement text. For example, it could be requested that a specific published paper describing the data set be referenced or that certain individuals who made exceptional contributions be acknowledged by investigators using the data set.

Recommendation: This field should be a variable length character field.

A.2 INVESTIGATOR INFORMATION

This section of the data set catalog contains information identifying the individuals, or group of individuals, who produced the data set. This normally means principal investigators and their immediate associates. Contact people and procedures are handled in a separate section. Separate roles have been defined for investigators that have been responsible for various aspects of the development of a data set. These roles reflect that EMAP monitoring has been completed using multiple teams of contractor and federal staff and that often no one investigator is responsible for all activities related to a data set. The fields may be repeated as many times as is necessary to represent multiple investigators. This section of the catalog references investigator names only; additional information about each named individual is provided in the Personnel Information section of the catalog (Section 3.13)

Princ Invest: Principal Investigator

Description: The principal investigator is the scientist responsible for all aspects of the steps associated with the creation of the data set including sampling design, sample collection, sample processing, data analysis, and publication. That person may be the technical director of an EMAP Resource Group, a lead scientist from one of the USEPA laboratories, an academic scientist working under a cooperative agreement, or a contractor.

Recommendation: The first name, middle initial, and last name of an individual should be given. Titles and other information should be omitted.

Sam Coll Invest: Sample Collection Investigator

Description: The individual who directly supervised the collection of samples should be named in this field.

Recommendation: The first name, middle initial, and last name of an individual should be given. Titles and other information should be omitted.

Sam Proc Invest: Sample Processing Investigator

Description: The individual who directly supervised the processing of samples should be named in this field.

Recommendation: The first name, middle initial, and last name of an individual should be given. Titles and other information should be omitted.

Data Anal Invest: Data Analysis Investigator

Description: The individual who directly supervised the analysis of data leading to the current data set should be named in this field.

Recommendation: The first name, middle initial, and last name of an individual should be given. Titles and other information should be omitted.

Add Invest: Additional Investigators:

Description: Other individuals associated with the creation of this data set should be named in this field.

Recommendation: The first name, middle initial, and last name of an individual should be given. Titles and other information should be omitted.

It is not necessary to provide information for principal investigators to complete a data set catalog entry. If principal investigator information is missing, the name of the contact person referenced in the directory will be given. If no contact person is identified, the name of the data center originating the data set will be given.

A.3 DATA SET ABSTRACT

A brief summary of the data set, like the abstract of a paper, allows a potential user to browse without learning all of the details. The General Keyword field can be repeated as many times as necessary. These fields are used to provide an overview of the data set as well as to provide information that can be extracted into indexes and summaries of aggregates of data sets.

Abstract: Abstract of the Data Set

Description: The Abstract should be a paragraph or two that summarizes the main points expanded in the following sections. That is, the short story on what data was collected, how it can be used, how it has been used, and how good is it.

Recommendation: This field is should be a variable length character field.

Gen Keyword: Keywords for the Data Set

Description: Any appropriate keywords should be listed in the Keyword field.

Recommendation: This field is restricted to no more than 40 characters. It may be repeated as often as necessary.

A.4 OBJECTIVES AND INTRODUCTION

This material can usually be abstracted from research proposals or project plans and provides the general information concerning why the particular study was completed.

The material presented is organized into several variable-length text fields. Information must be provided for each field to complete a data catalog entry. The Program Objective field will describe overall EMAP program objectives using standard text supplied by the Information Management Staff, and need not be completed by the catalog author.

Prog Objectiv: Program Objective

Description: The field presents the objective of the program under which the data set was collected. The objective of the program provides a general background for the specific data set. Many data sets will have the same program objective.

Recommendation: This field is a variable length text field. No limit is specified, but typically one to three paragraphs is sufficient to define the general objectives of a program.

Data Objectiv: Data Set Objective

Description: The objective or purpose of collecting the data set is specified in this field.

Recommendation: This field should be a variable length text field. No limit is specified, but typically one paragraph is sufficient to define the specific objectives of collecting a data set.

Data Backgrd: Data Set Background Information

Description: This field contains background information for the data set and helps the potential data user understand the rationale for collecting the data.that were to be met by the collection sets this data collection activity in a broader context of scientific motivation and interactions with other data available or to be collected.

Recommendation: This field should be a variable length text field. No limit is specified, but typically several paragraphs are used to provide background information for a data set.

Parameter Sum: Summary of data set parameters

Description: This field contains a summary of the parameters in the data set. The field is not a duplication of the data set dictionary and should not contain a simple list of parameters included in the data set.

Recommendation: This field should be a variable length text field. No limit is specified, but typically one or two paragraphs are used to summarize parameters in the data set.

A.5 METHODS

The methods employed to create the data set are summarized in this section of the catalog. More detailed descriptions of methods, including standard operating procedures (SOPs), are referenced in the data set catalog and are more appropriately stored in separately from the catalog.

Descriptions of methods are organized in two subsections: data acquisition (including sampling) and data preparation (including sample processing).

A.5.1 Data Acquisition

Samp Objectiv: Sampling Objective

Description: This field contains a brief summary of the objective of data collection or sampling.

Recommendation: This field is a variable length text field. No limit is specified, but typically one or two sentences is sufficient to define the general objectives of sampling.

Sampl Method: Sample Collection Methods Summary

Description: This field provides a brief summary of the sampling method used.

Recommendation: This field should be a variable length text field. No limit is specified, but typically one or two paragraphs is usually sufficient to summarize methods.

Beg Sampl Date: Beginning Sampling Date

Description: The earliest data acquisition date in the data set.

Recommendation: This field should be a date field.

End Sampl Date: Ending Sampling Date

Description: The latest data acquisition date in the data set.

Recommendation: This field should be a date field.

Platform: Sampling Platform

Description: This field contains information about the platform from which sampling was conducted. Platforms include boats, vehicles, satellites, etc.

Recommendation: This field could be restricted to specific terms defined in a look-up table; however, it is more flexible if formatted as a variable length text field, restricted to 80 characters.

Sample Equip: Sampling Equipment

Description: A description of the sampling equipment or instrument used to collect the samples is given in this field.

Recommendation: This field is a variable length text field. No limit is specified, but typically one or two sentences is sufficient to define the general objectives of sampling.

Equip Manufac: Manufacturer of Sampling Equipment

Description: This field contains the name of the manufacturer of the equipment specified in the field 'Sample Equip'.

Recommendation: This field is a variable length text field restricted to no more than 80 characters.

Key Variables: Key Variables

Description: This field presents those variables measured directly with the sampling equipment or instrument listed above. The field may have a value of "none" if no data were directly generated as a result of the use of particular equipment.

Recommendation: This field is a variable length text field having no length restrictions.

Sam Meth Cal: Sampling Method Calibration

Descriptions: Specific calibration procedures for sampling equipment or instrumentation is documented in this field.

Recommendation: This field should be a variable length text field.

Sam Qual Con: Sample Collection Quality Control

Descriptions: Specific procedures used to ensure the consistent quality of samples is documented in this field.

Recommendation: This field should be a variable length text field.

Sam Meth Ref: Sample Collection Method Reference

Description: This field contains the bibliographic reference to the complete description of the sample collection methods.

Recommendation: This field should be a variable length text field.

Note: The data set catalog should be linked to files containing the complete descriptions of methods and standard operating procedures. These text files may be extensive due to documentation of step-by-step procedures that are not normally provided in the data set catalog.

Sam Meth Dev: Sample Collection Method Deviations

Description: This field is used to document any known deviations from the methods referenced in the previous field. Departures from standard procedures need to be documented as completely as possible. These descriptions are useful to scientists interested in assessing the overall quality of the data. Further, these descriptions are necessary to evaluate the significance of any long-term changes identified from the integration of separate data sets collected in multiple years.

Recommendation: This field should be a variable length text field.

A.5.2 Data Preparation and Sample Processing

Proc Objectiv: Data Preparation Objective

Descriptions: This field contains a brief summary of the objective of data preparation and sample processing steps.

Recommendation: This field should be a variable length text field. No limit is specified, but typically one or two sentences is sufficient to define the general objectives of sampling.

Data Proc Sum: Data Processing Methods Summary

Descriptions: This field provides a brief summary of the methods used to process samples.

Recommendation: This field should be a variable length text field. No limit is specified, but typically one or two paragraphs is usually sufficient to describe processing methods.

Sampl Proc Calib: Sampling Processing Method Calibration

Description: Specific procedures used to calibrate instruments or gear used to process samples is documented in this field.

Recommendation: This field should be a variable length text field.

Proc Qual Con: Sample Processing Quality Control

Description:Specific procedures used to ensure the consistent quality of samples is documented in this field.

Recommendation: This field should be a variable length text field.

Samp Proc Ref: Sample Processing Method Reference

Description: This field contains the bibliographic reference to the complete description of the data processing and methods.

Recommendation: This field should be a variable length text field.

Note: The data set catalog should be linked to files containing the complete descriptions of methods and standard operating procedures. These text files may be extensive due to documentation of step-by-step procedures that are not normally provided in the data set catalog.

Sampl Proc Dev: Sample Processing Method Deviations

Description: This field is used to document any known deviations from the methods referenced in the previous field. Departures from standard procedures need to be documented as completely as possible. These descriptions will be useful to scientists interested in assessment the overall quality of the data and are necessary to assess long-term changes from the integration of data sets collected in multiple years. This field should be a variable length text field.

A.6 Data Manipulations

This section of the data set catalog provides documentation for any manipulations of the data subsequent to data acquisition and data preparation. Data manipulations include:

! conversions to different units (mg/l to mg/kg for example)

! transformation of continuous data values to discrete values (for example, conversion of benthic index values to nondegraded and degraded categories)

! deviation of values from existing data (for example, calculation of water density from temperature and salinity values)

Documentation of these data manipulations should be sufficiently detailed so that future users of the data will fully understand what was done. The fields used in this section may be repeated as many times as is necessary to document the data manipulations used to produce the final data set.

Deriv Value Name: Name of New or Modified Value

Description: This field contains the name of any new or modified value which was created as a result of data manipulations.

Recommendation: The length of this field should be restricted to 80 characters.

Data Manip Desc: Data Manipulation Description

Description: This field contains an in-depth description of any data manipulations used to create or modify the contents of the data set. This field is mandatory and should include complete descriptions of algorithms or reference information (e.g.

look up tables) applied. For complex formulas, remember that digital transfer between machines and software works best with simple ASCII text. It may be necessary to use a "pseudo-code" approach to define your algorithm in a stepwise fashion.

Recommendation: This field should be a variable length text field.

Data Manip Exmpl: Data Manipulation Examples

Description: This field contains a few examples of the input data and values derived from them using the algorithm described. Intermediate calculations need not be shown, but the examples should be sufficient for someone who works through or programs the algorithm to determine that they achieve the correct results.

Recommendation: This field should be a variable length text field.

Man Code File: Data Manipulation Computer Code File

Description: This field contains the name and location of the file containing the original computer code used to complete data manipulations. Providing an entry in this field is optional.

Recommendation: The length of this field should be restricted to 80 characters.

Man Code Lang: Data Manipulation Computer Code Language

Description: This field documents the language of the computer code used to complete data manipulations (for example, SAS, FORTRAN, BASIC, C, etc.). This field need should be completed if either the preceding or following items are used.

Recommendation: The length of this field should be restricted to 40 characters.

Manip Code: Data Manipulation Computer Code

Description: This field can be used to provide the specific segment of computer code that performs the calculations or derivations described in the Data Manip field. Completing this field is optional, but is encouraged if complex algorithms have been used.

Recommendation: This field should be a variable length text field.

A.7 Data Description

The data description contents and accuracy are primarily a function of the investigator submitting the data set. However, the description provides the basis for organizing the data itself in the EMAP relational data base management system (DBMS), and will be entered into the DBMS data dictionary and used heavily by the information management staff. This usage will generate feedback to the investigator and to this catalog document through efforts to standardize like parameter names and definitions across data sets, checks of maximum and minimum values, reports generating data record examples, and identification of related data sets.

A.7.1 Description of Parameters

Parameters in the data set are listed and described in this section of the data catalog. This information is abstracted from the data dictionary for the EMAP monitoring data base and may be presented in tabular form. The following information should be provided for each parameter in the data set:

Param Name: Parameter Name

Description: The name given to the quantity in the data set (e.g. a column name in the ORACLE data base and/or the ASCII text file.

Recommendation: This field should be a character field of length 40.

SAS Param Name: SAS Parameter Name

Description: This is an optional field to be used if the data set is distributed in a SAS data file format. It contains the shortened form (i.e. 8 character limit) of the parameter name used in SAS.

Recommendation: This field should be a character field of length 8.

Param Descrip: Parameter label or description

Description: The full English name of the quantity, with adjectives or modifiers as required to be scientifically precise.

Recommendation: This field should be a character field of length 80.

Units: Units of measurement

Description: The physical units of measure used for recording the value in the data set. Indexes and other unitless numerical values should be indicated with the term "Unitless". Qualitative values or quantities which are not associated with a measurement scale should be indicated "N/A"

Recommendation: This field should be a variable length text field.

Data Type: Parameter data type

Description: This field describes the fundamental digital storage type of the parameter, primarily for use with programs that read or interpret the data. For data in the relational data base, this will normally be numeric, character, or date. When an ASCII data file is being described, it may be useful to also specify numeric fields as integer, real, or complex. Data encoded in application specific formats should be described as binary.

Recommendation: This field should be a variable length text field.

Precision: Precision to which values are reported

Description: The precision of a value is the smallest value increment that is reported by the instrument or procedure by which the value is determined. For example, the last digit on an electronic display of an instrument represents the instrument precision. If the values in the data set are recorded or calculated at less than the instrument precision, then the actual smallest value increment in the data set should be entered in this field. For example, even though a GPS unit reports location to the nearest hundredth of a meter, the investigator decides it is most appropriate to report the data only to the nearest meter. Then the precision of the coordinate value would be 1 meter.

Recommendation: This field should be real numeric field, and should use the same units as the parameter value itself.

Accuracy: Accuracy of the data values

Description: The accuracy of a value is an indication of the total measurement error (random and/or systematic) associated with the value. This can be as small as +/- 1/2 of the instrument precision, but in most cases is significantly larger. The accuracy is normally determined by observing the deviations in calibration tests or repeated measurements of the same sample. Calculated values have accuracies determined from the individual accuracies by the standard statistical rules for propagation of error. The absolute value of the measured or calculated accuracy for the values of this parameter in the data set should be entered in this field.

Recommendation: This field should be a real numeric field and should use the same units as the parameter value itself.

Act Min Value: Minimum Value in Data Set

Description: This field is used to record the actual minimum value recorded in the data set. It is used for numeric fields only.

Recommendation: This field should be a numeric field.

Act Max Value: Maximum Value in Data Set

Description: This field is used to record the actual maximum value recorded in the data set. It is used for numeric fields only.

Recommendation: This field should be a numeric field.

A.7.2 Data Record Example

A display of a limited number of records or observations in the data set assists the data user in understanding the structure and composition of the data set. This section of the data set catalog presents in tabular form a display of several example records from the data set. For simple data sets, the first couple of records may suffice. If there is a complex structure to the data (widely varying parameters, or parameter dependent measurements, for example), it may be useful to select a few records that illustrate the range of values or the patterns of missing values. In no case should more than 20 records be included here. The records to be displayed can be provided by the investigator or can be generated from the EMAP relational data base after the data set is loaded into the data base. If the data set is complex or there are specific examples that should be shown, the investigator should elect to provide the material at the time the catalog document is written. The Example Data Values line is repeated once for each data set record included.

Special arrangements should be made if this field is to be used in the description of binary data sets (for example, SAS files, GIS coverages, or a satellite images). The field serves two purposes: to give the user a preview of the scientific data and to give the programmer who must access the data formatting and verification information. For the first purpose, if numbers are relevant, they can be extracted, and a "pseudo record" constructed. Similarly, a thumbnail version of a displayed GIS or image data set could be attached to the catalog document.

Actual Byte values can be included to meet the second purpose, if necessary - e.g. A TIFF image header and the corresponding values in the data file could be listed in the two fields.

Header Line: Column Names for Example Records

Description: This field gives the names of each column (in order) shown in the sample data records.

Recommendation: This field should be a variable length text field.

Exmpl Data Values: Example Data Records

Description: The example data values from a data set record are recorded in this field. They should be spaced so as to line up with the column names in the header line, when displayed in a fixed (non proportional) font.

Recommendation: This field should be a variable length text field.

A.7.3 Related Data Sets

The names and identification codes of data sets containing similar or related data are referenced in this section of the data catalog. The references should be entered exactly to make possible relational links (via the RDBMS) to the documentation for those data sets. These fields can be repeated to accommodate multiple data set references.

Related DS Name: Related Data Set Name

Description: The name of a data set containing similar or related data is recorded in this field. The contents of this field should match the contents of the "Data Set Name" field of the referenced data set.

Recommendation: The format of this field is the same as the Data Set Name field, and is specified in the guidance provided for building directory entries.

Related DS ID: Related Data Set Identification Code

Description: The identification code for a data set containing similar or related data is recorded in this field. The contents of this field should match the contents of the "Data Set Id" field of the referenced data set.

Recommendation: The format of this field is the same as the Data Set Id field, and is specified in the guidance provided for building directory entries.

A.8 GEOGRAPHIC AND SPATIAL INFORMATION

This section of the data set catalog will contain information about the spatial coverage of the data set. Information specific to the documentation of spatial data sets will also be provided in this section of the catalog. The documentation of geospatial data sets will be based upon the spatial metadata standards being developed by the Federal Geographic Data Committee (FGDC). Note, however, that the FGDC standards address only a subset of complete scientific data set documentation, and do so at a level of detail appropriate for a granule of data (i.e. a single GIS coverage or a single satellite image). The catalog contains descriptive information common to all data granules in a data set. Granule specific details in the FGDC standard are not appropriately covered in the catalog, but may be included in an inventory specifically designed for an individual data set.

The FGDC standards are intended to provide a common set of terminology and definitions; they explicitly do not specify (i) the means to organize information in a computer system, (ii) the means to organize information in a data transfer, or (iii) the means by which the information is transmitted , communicated, or presented to the user. Note also that the standards specify acceptable output formats for four types of values: (1) calendar dates (YYYYMMDD and variants thereof), (2) time of day (HHMMSSSS, either local, local with UT differential, or UT), (3) latitude and longitude (decimal degrees), (4) network addresses and file names (URL internet convention). The information need not be submitted or stored in these formats, although it may be convenient to do so, as long as the necessary information is present. Data management systems such as ORACLE include sophisticated date, time, numeric, and lexical conversion and manipulation functions that allow essentially arbitrary report formatting to be determined by the user.

Most of the relevant details of the sections of the FGDC metadata, i.e. data set identification, quality, parameters, and distribution information, will be addressed elsewhere in the catalog in the broader context of documenting all data sets uniformly. In the geographic coverage section, specific spatial data organization and reference information will be supplied if applicable to the data set.

Min Longitude: Minimum Longitude

Description: The bounding coordinates (West, East, North, South) give the limits of coverage of a data set expressed by latitude and longitude values. Under most circumstances the minimum longitude value is the Western Bounding Coordinate. For data sets that include a complete band of latitude around the Earth, the Western Bounding Coordinate shall be assigned the value -180.0, and the Eastern Bounding Coordinate shall be assigned the value 180.0. Latitude and longitude are expressed in decimal degrees.

Recommendation: This should be a real numeric field.

Max Longitude: Maximum Longitude

Description: The bounding coordinates (West, East, North, South) give the limits of coverage of a data set expressed by latitude and longitude values. Under most circumstances the maximum longitude value is the Eastern Bounding Coordinate. For data sets that include a complete band of latitude around the Earth, the Western Bounding Coordinate shall be assigned the value -180.0, and the Eastern Bounding Coordinate shall be assigned the value 180.0. Latitude and longitude are expressed in decimal degrees.

Recommendation: This should be a real numeric field.

Max Latitude: Maximum Latitude

Description: The bounding coordinates (West, East, North, South) give the limits of coverage of a data set expressed by latitude and longitude values. Under most circumstances, the maximum latitude will be the North Bounding Coordinate. Latitude and longitude are expressed in decimal degrees.

Recommendation: This should be a real numeric field.

Min Latitude: Minimum Latitude

Description: The bounding coordinates (West, East, North, South) give the limits of coverage of a data set expressed by latitude and longitude values. Under most circumstances the minimum latitude will be the South Bounding Coordinate. Latitude and longitude are expressed in decimal degrees.

Recommendation: This should be a real numeric field.

Geo Keyword: Name of the area or region

Description: This should be a searchable indirect spatial reference. An indirect spatial reference describes a location without using coordinates, usually by a feature such as a political entity (county, state), a road, or a geological province. The reference should uniquely identify the spatial extent of the data set, e.g. by using a name or a code that identifies the feature (such as a county FIPS code or a HUC code). To include multiple overlapping names (e.g. an EPA Region and the states within it), the Geo Keyword field can be repeated as many times as necessary.

Recommendation: This field should be a variable length character field.

Spatial Ref Meth: Direct Spatial Reference Method

Description: This field provides information on the system of objects used to represent space in the data set. Valid entries are limited to the values "Point", "Vector", or "Raster".

Recommendation: This should be a fixed character field of length 6.

Horiz Coord Sys: Horizontal Coordinate System Used

Description: This field names the reference frame or system from which linear or angular quantities are measured and assigned to the position that a point occupies. In general this will be "Geographic" (for latitude/longitude references) or the name of a planar map projection (such as Albers Equal Area; Universal Transverse Mercator; or Space Oblique Mercator). Where a special or unique projection is used (i.e. where the relationship between the coordinates and geographic coordinates is not known), use the term "Local System".

Recommendation: This should be a variable length character field

Horiz Resolution: Resolution of Horizontal Coordinates

Description: The minimum distance between two adjacent coordinate values or grid cells. In raster data sets, these values normally are the dimensions of the pixel or grid cell. In vector data sets, the resolution is the shortest line that is encoded in the data set.

Recommendation: This should be a real numeric field

Horiz Coord Units: Units for Horizontal Coordinates

Description: The units of measure in which the coordinates are reported.

Recommendation: This should be a variable length character field.

Vertical Coord Sys: Vertical Coordinate System

Description: This field names the reference frame or system from which vertical distances (altitudes or depths) are measured. This will normally consist of an Altitude Datum Name (e.g. North American Vertical Datum of 1988) or a Depth Datum Name (e.g. Local Surface, Mean Sea Level)

Recommendation: This field should be a variable length character field.

Vertical Resolution: Resolution of Vertical Coordinates

Description: The minimum vertical distance between two adjacent altitude or depth values.

Recommendation: This should be a real numeric field

Vertical Coord Units: Units for Vertical Coordinates

Description: The units of measure in which the vertical coordinates are reported.

Recommendation: This should be a variable length character field.

A.9 QUALITY CONTROL/QUALITY ASSURANCE

Quality control and quality assurance information is used to understand the limits of the data. This information includes: methods used to measure and ensure data quality, measurement quality objectives, and summaries of data quality parameters. Specific data collected to assess data quality may be included in this section of the data set catalog, or may be included in a separate data set that is referenced in the data set catalog.

Meas Qual Obj: Measurement Quality Objectives

Description: This field lists any specific a priori objectives established for measurement or sampling data.

Recommendation: This should be a variable length character field.

QA/QC Meth: Quality Assurance/Control Methods

Description: This field should hold a brief description of the methods used to measure and ensure data quality.

Recommendation: This should be a variable length character field.

Act Meas Quality: Actual Measurement Quality

Description: The results of any assessments of measurement quality, including the measurement error for parameters and variables and a summary of data quality parameters. For spatial data, the positional accuracy should be noted.

Recommendation: This should be a variable length character field.

Sources of Error: Sources of Error

Description: Any uncontrolled factors which may have impacted the quality of the measurements should be described in this field. In particular, sources of error that cause the actual measurement quality to vary greatly from the measurement quality objective for a parameter should be noted. Also include systematic measurement errors that may not be reflected in quality assessments, but may affect the usefulness of the data in analysis.

Recommendation: This should be a variable length character field.

Known Data Prob: Known Problems with the Data

Description: This field should contain a discussion of problems that have been documented for the data set.

Recommendation: This should be a variable length character field.

Conf Level Stmnt: Confidence Level/Accuracy Judgement

Description: Subjective statement of Investigator's confidence in the data.

Recommendation: This should be a variable length character field.

Allow Min Value: Allowable Minimum Values

Description: List the parameter name and the allowable (physically, biological, mathematical limits) maximum value. Repeat the field as many times as required to describe the parameters in the data set.

Recommendation: This should be a variable length character field.

Allow Max Value: Allowable Maximum Values

Description: List the parameter name and the allowable (physically, biological, mathematical limits) minimum value. Repeat the field as many times as required to describe the parameters in the data set

Recommendation: This should be a variable length character field.

QA Ref Data: QA Reference Data

Description: May include actual reference data or the name of a file or document that has the appropriate reference data.

Recommendation: This should be a variable length character field.

A.10 DATA ACCESS

In an integrated environment, metadata for a data set will be directly linked to the data. Data access will be direct and the user will require no additional information. However, there will be instances when direct links from the metadata to the data will not be possible. In those instance where the metadata are not linked to data, it is necessary to provide information regarding data access.

This section is currently designed to provide users with information on how to access data. The section may be expanded to include information concerning data archiving; however, data archiving is primarily a data management function that is of little interest to the user. It is recommended that data archiving information be stored separately from the data set catalog.

Data Access: Data Access Procedures

Description: Procedures for accessing data is given in this field. Procedures should provide general information about the different ways users can access data including telephone contact, anonymous FTP WWW, Gopher, WAIS, dial-in lines, etc.

Recommendation: This field should be a variable length text field.

Data Access Restrict: Data Access Restrictions

Description: If there are restrictions on accessing or using the data, they should be explained clearly in this field.

Recommendation: This field should be a variable length text field.

Data Access Contact: Data Access Contact person

Description: The primary person to contact to obtain the data set or information about the data set. For a data set that is located in a Task Group data base, this will normally be the data librarian. For data sets available through the EMAP IMS, the IMS staff will provide the name of the appropriate contact. The name and address of the person or organization to contact should be included in this field using the fields that provide information for the principal investigator as a guide.

Recommendation: This field should be a variable length text field.

Data Set Format: Data Set Format

Description: Give a description of the format of the file(s) containing the data set, e.g. ASCII, SAS, ARC/Info Export. Repeat this field once for every format type that is available.

Recommendation: This field should be a variable length text field.

FTP Infor: Information Concerning Anonymous FTP

Description: Information concerning the access of EMAP data sets using the Internet and anonymous file transfer protocols should be provided as part of the data access documentation. This information should include the URL (Universal Resource Locator), name of the node on the network, login and password information, and the name of the directories containing the data of interest. This information should be provided for those data sets that contain unrestricted information only.

Recommendation: This field should be a variable length text field.

Gopher Infor: Information Concerning Gopher

Description: Information concerning the access of EMAP data sets using Gopher services via the Internet should be provided as part of the data access documentation. This information should include the URL (Universal Resource Locator), name of the node on the network, login and password information, and the name of the directories containing the data of interest. This information should be provided for those data sets that contain unrestricted information only.

Recommendation: This field should be a variable length text field.

WWW Infor: Information Concerning World Wide Web

Description: Information concerning the access of EMAP data sets using WWW browsers via the Internet should be provided as part of the data access documentation. This information should include the URL (Universal Resource Locator), name of the node on the network, login and password information, and the name of the directories containing the data of interest. This information should be provided for those data sets that contain unrestricted information only.

Recommendation: This field should be a variable length text field.

EMAP CD-ROM: EMAP CD-ROM Containing the Data set

Description: CD-ROM is becoming a widely used method to distribute data and metadata. It is likely that EMAP will also distribute data using CD-ROM technology sometime in the near future. This field documents the name of the CD-ROM that contains the data set. The field may be repeated in where the data set is saved on more than one CD-ROM.

Recommendation: This field should be a variable length text field.

A.11 REFERENCES

This section of the data set catalog provides a list of any published documentation relevant to the data collected. Documentation may include manufacturer's instruction manuals, government technical manuals, user's guides, etc. Also referenced should be any technical reports and scientific publications concerning the methods, instruments, or data described in this document. Publications by the Principal Investigator or investigating group that would help a reader understand or analyze the data are particularly important. The format of the bibliographic references is taken from the EMAP bibliographic data base specification. The section is broken into two subsections to accommodate references that will be maintained in the EMAP bibliographic data base (detailed reference format) and those background and general references which are primarily of use to the reader and will not be tracked (brief reference format)

A.11.1 EMAP References

References given in this section are those published or used as part of EMAP and which should be tracked in the EMAP bibliographic data base.

Ref Type: Reference Type

Description: This field indicates the type of documentation being referenced. Valid values include Journal Article, Workshop Proceedings, Book, Report, Film, Video Tape, Audio Tape, CD-ROM.

Recommendation: This should be a variable length character field.

Ref Author: Reference Author

Description: Names of the authors are given in this field. It is repeated (in conjunction with the following field) as many times as necessary to list all authors.

Recommendation: This should be a variable length character field.

Ref Authors Affil: Reference Author's Affiliation

Description: The professional affiliation and address of each author is provided in this field. It is paired with the previous field, and is repeated once for each author.

Recommendation: This should be a variable length character field.

Ref Title: Title of Reference

Description: The full title of the referenced material is given in this field.

Recommendation: This should be a variable length character field.

Volume Title: Journal or Volume Title

Description: If the reference is an article in a journal, a proceedings volume, or other compendium, the title of the full work is given in this field.

Recommendation: This should be a variable length character field.

Volume Editor: Journal or Volume Editor

Description: If the reference is an article in a compendium or other edited work, the name of the overall editor is entered in this field.

Recommendation: This should be a variable length character field.

Page Ref: Page and Volume Reference

Description: The standard volume and page citation information should be entered into this field.

Recommendation: This should be a variable length character field.

Date of Ref: Date the Reference was Published

Description: This field contains the publication date for the reference.

Recommendation: This should be a variable length character field.

Place of Pub: Location of Publishing Organization

Description: The location (city, state, country) of the organization publishing the reference is listed in this field.

Recommendation: This should be a variable length character field.

Publisher: Name of Publishing Organization

Description: The business name of the publishing organization is entered in this field.

Recommendation: This should be a variable length character field.

Ref Other ID: Reference Report Number or Other ID

Description: For special reports, technical memoranda, or other material that carries an internal organizational identification code or report number, it should be recorded in this field.

Recommendation: This should be a variable length character field.

Procite Rec Num: Procite Record Number for the Reference

Description: If the reference has a Procite record number, it should be entered in this field.

Recommendation: This field should be formatted to match the Procite record number format.

A.11.2 Background References

The references listed in this section are materials in the open literature which may support the collection, use, and interpretation of the data set, but are not directly related to or products of EMAP. As such they need not be tracked in the EMAP bibliographic data base, but should be captured as part of the documentation. A single text field is provided, but it should contain all of the relevant information that would otherwise have been listed in the previous section.

Supt Ref: Supporting Reference

Description: This field contains the reference information as described above for a single supporting reference relevant to the data set. It may be repeated as many times as required.

Recommendation: This field should be a variable length text field.

A.12 GLOSSARY AND TABLE OF ACRONYMS

The detailed documentation for each data set is likely to contain terms and acronyms that are unfamiliar to some potential users of the data. These acronyms will be defined the first time they are used; however, due to the length of the documentation, a separate table of acronyms is suggested to assist the user.

Acronym: Acronym used in the Detailed Documentation

Description: Acronyms used in the text of the detailed documentation should be listed using this field. The field should be repeated once for each acronym, in conjunction with the following field.

Recommendation: This field should be a variable length text field.

Acronym Def: Definition of Acronym

Description: The full expanded version of each acronym should be given in this field. It is paired with the preceding field.

Recommendation: This field should be a variable length text field.

A.13 PERSONNEL INFORMATION

This section of the data set catalog contains information identifying the individuals who are associated with the data set and named in the data set catalog. This normally includes principal investigators, co-investigators, catalog authors and contributors, data librarians, and other contact people. The fields used to identify data set personnel are similar to those used in the data set contact fields of the data set directory and may be linked to those fields. These fields may be repeated as many times as is necessary to identify people.

The personnel named in the catalog entry and described in this section should be identified by their role with respect to this data set using the Role field. For example, if the person is named as Cat Author in section 1, then Role in this section would be Catalog Author. Similarly, the role is used to identify the Principle Investigator, the sub-investigators, and the contact people. There may be multiple individuals listed for a role or multiple roles for an individual.

In implementation, these fields may be used to reconstruct the name fields in the primary sections if desired. That is, the Role, Last Name, First Name, and Middle Initial fields may be selected and concatenated to form the Cat Author, Principal Investigator Contact Person, etc., fields. However, we recommend that these fields be written and maintained separately within the catalog section of the data base, because the role field of the contacts data objects will be frequently updated and easily corrupted. Once lost in a blanket update of the contacts data object, catalog specific information may not be recoverable, especially if the catalog entry is not used for QA'd for an appreciable period after the update.

Title: Formal title

Description: The formal title for investigators is given in this field (i.e., Mr., Mrs., Miss, Dr., or Ms).

Recommendation: This field is restricted to no more than 5 characters. Valid entries are given in Table A-2.

Table A-2. Valid entries for contact title

Dr	Miss	Mr
Miss	Prof	

Lst Name: Last name

Recommendation: This field is restricted to no more than 30 characters.

Frst Name: First name

Recommendation: This field is restricted to no more than 15 characters.

Mid Init: Middle Initial

Recommendation: This field is restricted to no more than 1 character.

Role: Role described in the Catalog

Description: The role is the name of the field in catalog sections 1-12 in which the person is listed, e.g., Catalog Author.

Recommendation: This is a variable length character field. Valid entries role are provided in Table A-3.

Table A-3. Valid entries for role
Director, EMAP
Technical Director
Technical Coordinator
Task Group Information Manager
Data Center Information Manager
Data Base Administrator
Data Librarian
Regional Environmental Services Division Director
Principal Investigator
Sample Collection Investigator
Sample Processing Investigator
Data Analysis Investigator
Additional Investigator
Catalog Author
Reference Author
Quality Assurance Officer
Chief Scientist
Project Manager

Address1: Line 1 of address
Address2: Line 2 of address
Address3: Line 3 of address
Address4: Line 4 of address

Description: Organizational names, street addresses, rural route codes, or post office box numbers may be specified in four address fields (Address1 through Address4).

Recommendation: The four address fields are restricted to no more than 40 characters each.

City: City

Recommendation: This field is restricted to no more than 30 characters.

State: State

Recommendation: This field is restricted to no more than 2 characters. The two-letter state abbreviation is used to identify each state.

Zip: Zip code

Recommendation: This field is restricted to 10 characters to identify the zip code.

Country: Country

Recommendation: This field is restricted to no more than 40 characters.

Voice Phone: Voice phone number

Recommendation: Phone numbers should include area codes. Up to 18 characters can be used. Phone extension numbers, if applicable, can be added at the end of the phone numbers as 'X1234'.

FAX Phone: FAX phone number

Recommendation: Phone numbers should include area codes. Up to 18 characters can be used. Phone extension numbers, if applicable, can be added at the end of the phone numbers as 'X1234'.

Fields specifying the electronic mail addresses may be repeated as many times as needed to reflect multiple electronic mail addresses. Email addresses should be given for both internal (EPA) and external (Internet) networks to be accessible to the widest range of potential users.

Email Address: Email address

Recommendation: This field is restricted to no more than 80 characters.

Email Network: Email network

Recommendation: This field is restricted to no more than 20 characters. Valid entries for this field are given in Table A-4.

Table A-4. Valid names for email networks
Bitnet
Internet
OMNET
Telemail
USEPA All-in-One
USEPA VAX

Add EM Info: Additional Email Information

Description: Any additional information concerning electronic mail addresses may be given in this field. For example, this field may be used to describe which Email network is preferred by a particular person or organization.

Recommendation: This field is restricted to no more than 80 characters.