

APPENDIX K

DRAFT

DATA MANAGEMENT PLAN

Appalachian Highlands Inventory and Monitoring Network

December 2004

**Appalachian Highlands Inventory and Monitoring Network
National Park Service
Blue Ridge Parkway
199 Hemphill Knob Rd.
Asheville, NC 28801**

ACKNOWLEDGMENTS: The APHN staff would like to thank many people for their contributions to our data management program and the writing of this document, especially Debbie Angel, Dorothy Mortenson, Margaret Beer, John Boetch, Christina Wright , and the members of the Data Management Planning Working Group, who contributed to the chapters listed below:

<u>Name</u>	<u>Network</u>	<u>Chapter</u>
John Boetsch	NCCN	II, III, and IV
Rob Daley	GRYN	III
Patrick Flaherty	APHN	IV
Geoff Sanders	NCRN	V
Debbie Angell	SODN	VI
Sara Stevens	NCBN	IX
Margaret Beer	NCPN	X
Karen Oakley	USGS	VIII

Table of Contents

Acknowledgements.....i

Chapters

1. Introduction1

 1.1 I&M Program Overview1

 1.2 APHN Organization and Management1

 1.3 APHN Data Management Goal5

 1.4 Scope of the Data Management Plan6

2. Data Management Roles and Responsibilities8

 2.1 Data Stewardship8

 2.2 Data Management Goals9

 2.3 Data Management Coordination10

3. Project Work Flow and the Data Management Process.....12

 3.1 Project Workflow12

4. Data Management Resources.....17

 4.1 Computer Resources Infrastructure17

 4.2 Network Systems Architecture19

 4.3 National Information Management Systems21

 4.4 Other National-Level Data Management Tools.....23

5. Data Acquisition and Processing25

 5.1 Programmatic Data25

 5.2 Non-Programmatic Data29

6. Quality Assurance/Quality Control.....32

 6.1 NPS Mandate for Data Quality32

 6.2 QA/QC Mechanisms33

 6.3 Roles and Responsibilities34

 6.4 Goals and Objectives35

 6.5 Data Collection35

 6.6 Data Entry38

 6.6.1 Methods for Reducing Collection Errors38

 6.7 Verification and Validation Procedures40

 6.8 Version control.....43

 6.8.1 Version Control Options44

 6.9 Data Quality Review and Communication45

7. Data Documentation/Metadata48

7.1 Purpose of Metadata 48
 7.2 NPS Integrated System 48
 7.3 Metadata Process/Workflow 49
 8. Data Summary and Reporting 52
 8.1 Data Analysis 52
 8.2. Reporting of Monitoring Information 53
 9. Data Dissemination 56
 9.1 Data Ownership 56
 9.2 Data Distribution 56
 9.3 Handling Sensitive Data 58
 10. Data Maintenance, Storage and Archiving 61
 10.1 Digital Data Maintenance 61
 10.2 Storage and Archiving Procedures- Digital Data 62
 10.3 Storage and Archiving Procedures- Documents and Objects 63
 11. Literature Cited 66

Figures

1.1 APHN Network Map 4
 3.1 Workflow Diagram for Generic I&M Project 13
 3.2 Data Processing Steps 16
 4.1 APHN Computer Resources 18
 4.2 Vital Signs Directory Structure 18
 4.3 Linkage to Common Tables 20
 4.4 National-Level Architecture 22
 7.1 NPS Integrated Metadata System 50

Tables

1.1 Categories of Data Products 5
 2.1 APHN Data Management Staff Resources 20
 4.1 Groupings for Common Lookup Tables 53
 7.1 Reporting of Monitoring Data 54
 9.1 Data Types Uploaded to Web Applications 57

Standard Operating Procedures (SOPs)

Standard Operating Procedure 1: Data Dictionary 67
 Standard Operating Procedure 2: Database Specifications 69
 Standard Operating Procedure 3: GIS Specifications 72
 Standard Operating Procedure 4: GPS Specifications 75
 Standard Operating Procedure 5: Archiving, naming, storage, versioning 78

Appendices

Appendix 1: NPspecies List of Editing Capabilities 78

Appendix 2: APHN Software and Computer Resources	79
--	----

Executive Summary

The central mission of the NPS Inventory and Monitoring Program is to provide timely and usable scientific information about the status and trends of park resources to park managers. Fundamental to this mission is a level of confidence in the information that supports management decisions. Data management is the combination of attitudes, habits, procedures, standards, and infrastructure related to the creation, maintenance and disposition of data and its supporting information. Seen in this way, it becomes obvious that data management is the key to success of any long-term monitoring initiative. Proper data management is the means by which a thorough understanding of the scope, origin, and value of scientific information about our natural resources can become a part of the Park Service heritage.

The goal of our data management program is to ensure the quality, interpretability, security, longevity and availability of ecological data and related information resulting from resource inventory and monitoring efforts:

- *Quality* – Awareness of the value of information and its underlying data is fundamental to its proper use. Our objective is to ensure that appropriate quality assurance measures are taken during all phases of project development, data acquisition, data handling, summary and analysis, reporting, and archival. These should reflect current best practices and meet rigorous scientific standards. Standards and procedures can only accomplish so much, therefore an important part of quality assurance is to continually encourage careful attitudes and good habits among all staff involved in creating, collecting, handling, and interpreting data.
- *Interpretability* – A data set is only useful if it can be readily understood and appropriately interpreted in the context of its original scope and intent. Data taken out of context can lead to misinterpretation, misunderstanding, and bad management decisions. Similarly, data sets that are obscure, complex or poorly documented can be easily misused. Sufficient documentation should accompany each data set and any reports and summaries derived from it, so as to ensure that users will have an informed appreciation of its applicability and limitations.
- *Security* – Our objective is to make certain that both digital and analog forms of source data are maintained and archived in an environment that provides appropriate levels of access to project managers, technicians, decision makers, and others. Our data management program will take advantage of existing systems for network security and systems backup, and augment these with specific measures aimed at ensuring the long-term security and integrity of our data.
- *Longevity* – Countless data sets have been lost over time simply because they were not sufficiently documented and organized when they were created. Too

often data are left in a condition that renders them effectively irretrievable – either because the format is outdated (e.g., punch cards that can no longer be read), or more often because there is not enough documentation to inform subsequent users of the data collection methods, the scope and intent of the data set, the quality assurance procedures, the data format, or the intended use. Without sufficient information about a data set we lose confidence in its quality and applicability, which leaves it useless and unused. The longevity of a data set can be enhanced by thorough documentation, by maintaining the data in a widely interpretable format, and by appropriate archival measures. Although this requires an initial investment of time and effort upon creation of the data set, this investment almost certainly pays off over time because the dataset is much more likely to be used if it is properly documented and formatted. Furthermore, simply using a dataset enhances its longevity because its value is being realized and enhanced through use.

- *Availability* – Natural resource information can only be useful for informing decisions if it is available to managers at the right time and in a usable form. Our objective is to expand the availability of natural resource information by ensuring that the products of inventory and monitoring efforts are created, documented and maintained in a manner that is transparent to the potential users of these products.

Although our primary responsibility is for ecological data and related information derived from Inventory and Monitoring Program activities, it is anticipated that the scope of our data management efforts will expand to include all significant natural resource data specific to the Appalachian Highlands Network and the parks of which it is composed.

Things to strive for in data management:

- ... confidence among users in the stability and security of all forms of information
- ... easy access to general information, and appropriate safeguards for sensitive information
- ... continual awareness of the scope, usability, and limitations of each data set
- ... infrastructure and documentation that encourages data exploration by a wide array of users
- ... compatibility of data sets for interdisciplinary data integration and analysis
- ... implementation and acceptance of standards and procedures that direct the flow of information – from project inception to its conclusion, from data collection through reporting and archival – and that ensure appropriate documentation and data quality
- ... a proper balance between the structure needed to ensure quality and usability, and the flexibility needed to meet specific needs and allow for new designs