Data and Information Management Plan

Version 1.0
September 2005

National Park Service
Greater Yellowstone Inventory and Monitoring Network

Bighorn Canyon National Recreation Area
Grand Teton National Park
John D. Rockefeller Jr. Memorial Parkway
Yellowstone National Park

NPS GRYN Offices
USGS Northern Rocky Mountain Science Center
229 AJM Johnson
Montana State University
Bozeman, MT 59717

Prepared by

Rob Daley
GRYN Data Manager

With contributions from
Sara Stevens, Margaret Beer, Deborah Angell, Dorothy Mortenson, Chad Jacobson, Ann Rodman, Richard Easterbrook, Cassity Bromley, Shannon Savage, Doug Wilder, Gareth Rowell, Michael Williams, Geoffrey Sanders, Teresa Leibfreid, Bill Moore, Patrick Flaherty, Fred Dieffenbach, Lisa Nelson, Velma Potash, Gary Entsminger, Joe Gregson and Steve Fancy

Web Site (subject to change):
http://www.nature.nps.gov/im/units/gryn/index.shtml

Recommended citation:
Daley R. 2005. Data and Information Management Plan, Greater Yellowstone Inventory and Monitoring Network. Bozeman, MT: National Park Service, Greater Yellowstone Network. 82 p. plus appendices.

Table of Contents

LIST OF FIGURES ..... **IV**

LIST OF TABLES ..... **IV**

LIST OF APPENDICES ..... **IV**

EXECUTIVE SUMMARY ..... **V**

1. INTRODUCTION **1**

1.1 GRYN ORGANIZATION 2
1.2 DATA AND INFORMATION SYNTHESIS 2
1.3 GOALS AND OBJECTIVES FOR MANAGING NATURAL RESOURCE INFORMATION ASSETS 2
1.4 DATA MANAGEMENT PLAN SCOPE AND REVISION SCHEDULE 3

2. DATA STEWARDSHIP ROLES AND RESPONSIBILITIES **4**

2.1 DEFINITIONS OF ROLE AND RESPONSIBILITY 5
2.2 PRIMARY ROLES AND RESPONSIBILITIES FOR DATA MANAGEMENT 5
2.3 COORDINATION WITH PARKS AND OTHERS 7
2.3.1 Network of Networks *7*
2.3.2 External Coordination *8*
2.4 DOCUMENTATION IS KEY 8
2.5 PROJECT STEWARDSHIP 9
2.6 DATA STEWARDSHIP 9
2.6.1 Awareness and Accountability *9*
2.6.2 The Hub of Data Stewardship *10*

3. DATA MANAGEMENT PROCESS AND WORK FLOW **13**

3.1 PROJECT WORK FLOW 13
3.2 DATA LIFE CYCLE 15
3.3 INTEGRATING AND SHARING DATA PRODUCTS 18
3.3.1 Data distribution *18*

4. DATA MANAGEMENT INFRASTRUCTURE AND APPLICATIONS **20**

4.1 NETWORK COMPUTER INFRASTRUCTURE 20
4.2 NETWORK INFORMATION MANAGEMENT SYSTEMS 20
4.3 INFORMATION SYSTEM INFRASTRUCTURE 21
4.4 NATIONAL LEVEL I&M DATABASE APPLICATIONS 21
4.5 PROJECT DATABASE STANDARDS 24
4.5.1 Levels of data standards *25*
4.5.2 Project tracking application *25*

5. DATA ACQUISITION AND PROCESSING **27**

5.1 DATABASE DESIGN 28
5.2 DATA ACQUISITION 28
5.2.1 Data Discovery and Data Mining *28*
5.2.2 Field Studies *30*
5.3 DATA ENTRY 32
5.4 CHANGES TO DATA COLLECTION PROCEDURES/PROTOCOLS 33
5.5 DATA SOURCES 33
5.5.1 Park Data *33*
5.5.2 Regional and National Programs *34*
5.5.3 Processing National Park Service Data *34*
5.5.4 External Data *36*
5.5.5 Processing External Data *36*

6. DATA QUALITY ASSURANCE **38**
6.1 NPS MANDATE FOR QUALITY 38
6.2 DEFINITIONS OF QUALITY ASSURANCE AND QUALITY CONTROL 39
6.3 ROLES AND RESPONSIBILITIES FOR QUALITY ASSURANCE 40
6.4 GOALS AND OBJECTIVES FOR QUALITY ASSURANCE 41
6.5 DATA COLLECTION 42
6.5.1 Suggested Methods to reduce errors during field data collection *42*
6.6 QUALITY ASSURANCE FOR DATA ENTRY 44
6.6.1 Suggested Methods to reduce errors during data entry *45*
6.7 VERIFICATION AND VALIDATION PROCEDURES 46
6.7.1 Suggested Methods for Data Verification *47*
6.7.2 Data Validation *48*
6.7.3 Suggested Methods for Data Validation *49*
6.8 VERSION CONTROL 51
6.9 DATA QUALITY REVIEW AND COMMUNICATION 52

7. DATA DOCUMENTATION **56**
7.1 PURPOSE OF METADATA 56
7.2 NPS INTEGRATED METADATA SYSTEM PLAN AND TOOLS 56
7.3 METADATA PROCESS AND WORKFLOW 58

8. DATA MANAGEMENT SUPPORT FOR ANALYSIS AND REPORTING **61**
8.1 PERIODIC AND ANNUAL REPORTING 61
8.2 LONG-TERM TRENDS ANALYSIS AND REPORTING 61

9. DATA DISTRIBUTION **63**
9.1 NATIONAL PARK SERVICE POLICY ON DATA OWNERSHIP 63
9.2 ESTABLISHING DATA OWNERSHIP GUIDELINES 64
9.3 DATA DISTRIBUTION 65
9.3.1 Data Distribution Mechanisms *65*
9.4 DATA CLASSIFICATION: PROTECTED VS. PUBLIC 67
9.5 ACCESS RESTRICTIONS ON SENSITIVE DATA 69
9.6 NPS INVENTORY AND MONITORING DATA AVAILABILITY 70
9.7 DATA ACCESS AND REQUEST PROCEDURES 71
9.8 DATA QUALITY FEEDBACK 71

10. DATA MAINTENANCE, STORAGE AND ARCHIVING **72**
10.1 DIGITAL DATA MAINTENANCE 72
10.1.1 Short-term data sets *72*
10.1.2 Long-term monitoring data sets *73*
10.2 QUALITY CONTROL FOR CONVERTED DATA 73
10.3 STORAGE AND ARCHIVING PROCEDURES FOR DIGITAL DATA 74
10.4 DIRECTORY STRUCTURE FOR INDIVIDUAL PROJECTS 74
10.5 BACKUP PROCEDURES FOR DIGITAL DATA 75
10.6 DATA AND NETWORK SECURITY 75
10.7 STORAGE AND ARCHIVING PROCEDURES FOR DOCUMENTS AND PHYSICAL OBJECTS 75
10.7.1 Documents *76*
10.7.2 Specimens *77*
10.7.3 Photographs *77*
10.8 ROLE OF CURATORS IN STORAGE AND ARCHIVING PROCEDURES 77

CHANGE PROCEDURES AND REVISION HISTORY LOG **78**

REFERENCES **79**

List of Figures

Figure 1.1. National Park Service Units served by the GRYN. 1
Figure 2.1. Joint Project and Data Management Responsibilities 11
Figure 3.1. Conceptual model of project work flow 14
Figure 3.2. Diagram of the typical project data life cycle. 17
Figure 3.3. Storing and disseminating project information. 18
Figure 3.4. Steps involved in product distribution. 19
Figure 4.1. Network Computer Architecture 20
Figure 4.2. Model of the national-level application architecture. 22
Figure 4.3. Water Quality Data Flow 23
Figure 4.4. Common lookup tables and satellite databases 24
Figure 4.5. Different levels of data standards 25
Figure 5.1. Natural Resource Data Management Framework 35
Figure 6.1. QA/QC controls applied at progressive stages of a project 40
Figure 7.1. NPS Integrated Metadata System Plan 57

List of Tables

TABLE 2.1. CATEGORIES OF DATA STEWARDSHIP INVOLVING ALL NETWORK PERSONNEL 4
TABLE 2.2. SUMMARY OF ROLES AND RESPONSIBILITIES 6
TABLE 3.1. REPOSITORIES FOR GRYN PRODUCTS 18
TABLE 5.1. ABBREVIATED DATA DEVELOPMENT MODEL 27
TABLE 6.1. CHECKLIST OF FUNDAMENTAL QUALITY CONTROL ACTIVITIES FOR A PROJECT 55
TABLE 9.1. ONLINE DATABASE APPLICATIONS AND RELATED DATA TYPES. 66

List of Appendices

Appendix A: Roles, Responsibilities, and Awareness Levels for Data Management vii
Appendix B: GRYN Folder and File Naming Conventions xv
Appendix C: Freedom of Information Act (FOIA) and Sensitive Data xxvii
Appendix D: Using the Global Positioning System and Portable Data Recorders xxxv

Executive Summary

Established in 2000 as one of 32 field offices in the NPS Natural Resource Inventory and Monitoring Program, the Greater Yellowstone Network (GRYN) provides coordination and services for the inventory and long term monitoring of selected natural resources in Bighorn Canyon National Recreation Area, Grand Teton National Park (including John D. Rockefeller Jr. Memorial Parkway), and Yellowstone National Park. As a component of the network's Vital Signs Monitoring Plan this data management plan outlines the strategy and guidelines for thorough, integrated, and coordinated resource information management activities to meet NPS-wide information requirements and park level inventory and monitoring information needs. The content of this plan includes overview descriptions, guiding principles, important concepts, and elements of the Service's evolving enterprise architecture for its natural resource line of business. Specific procedures and best practices not covered in this plan are found in the data management and other sections of the network's individual monitoring protocols.

About one-third of I&M program and network resources are devoted to generating, managing, and sharing data and information resources that are well planned, highly organized, easily accessible, completely documented, and scientifically credible. The purpose of emphasizing data management is to provide stable, long-term support for natural resource managers and decision makers in the NPS, as well as non-NPS partners, scientists, and other stakeholders.

The foundation for effective data management begins with specifying data stewardship roles and responsibilities for every aspect of program and project planning and operations. This plan lists several roles and associated responsibilities for data management that cover data producers through end users. Management level support for these responsibilities, particularly at parks, is required to meet I&M program goals. Park managers need to plan for and provide funds and time to adequately manage data and information resources generated by I&M program and other natural resource management activities. Managers at all levels also need to provide accountability for data management responsibilities as performance elements for all natural resource positions. Emphasis is placed on the collaborative nature of successful data stewardship, and the network promotes a teamwork approach to data management. Because it is essential to establish, communicate, and support data stewardship responsibilities for all staff at all stages of a project, chapter three discusses project workflow and data management issues and opportunities. A standard data lifecycle model helps demonstrate the relationship between park-level data collection for inventory and monitoring projects with NPS service-wide databases and data distribution services.

The network operates and maintains a NPS-compliant information technology infrastructure at offices currently located on the Montana State University campus in Bozeman. Hardware, software, security provisions, and local area network services are managed by GRYN staff with support from regional and park level technology service staff. The network maintains project-specific database applications and data sets designed and populated based on inventory study plans and specific monitoring objectives, often in partnership with non-NPS partners such as the US Geological Survey and the US Forest Service. The data stored in these local applications are transferred at appropriate intervals to NPS service-wide

databases and repositories including NPSpecies, NatureBib, NPSTORET, the NPS Natural Resource GIS Metadata and Data Clearinghouse, and the NPS Biodiversity Data Store.

The network necessarily focuses on collecting and managing data that are required to meet monitoring objectives specified in the vital sign monitoring protocols. GRYN also works with staff from network parks, the NPS regional and Washington offices, and other partner entities to identify, manage, and integrate legacy data from past monitoring efforts and relevant data sources from ongoing monitoring that are not based strictly on protocols generated by the NPS I&M program. Chapter five discusses the networks options for recording data in the field, its approach to database design and data discovery, acquisition, and entry, and the treatment of data from different organizational levels in the NPS.

An important data management goal discussed in chapter six is how to ensure that I&M projects produce data that meet quality requirements specified in vital sign monitoring protocols. The plan discusses ways to reduce errors when entering data, procedures for verifying and validating data, and methods to share information about data quality to promote the appropriate application and use of data resources. Another principal component of effective data management operations is data documentation, discussed in the context of network activities involving NPS metadata creation and distribution tools in chapter seven. Chapter eight follows with a brief discussion of the supporting role of data management for analysis and reporting.

GRYN is committed to making all available and relevant data easily accessible while providing for the appropriate level of protection for sensitive data resources. Procedures and solutions for making data available to park natural resource managers and decision makers, as well as other stakeholders and interested parties are presented in Chapter nine. In the final chapter the long term elements of data maintenance, storage, and archiving are addressed for electronic as well as physical (specimen) and hardcopy data and information resources.

Current and future GIS and information management strategies and solutions at each park represent valuable information about what works and what needs improvement in the realm of data management. Staff from the network and the parks will coordinate and integrate data management activities based on common tools, best practices and shared objectives. This plan is a key mechanism to help understand and address the relationships between park level information needs supported by park GIS and data management activities, and I&M program organization and standards for Network data management.