A Future For Data: An Overview of Data Management for Analysis, Decision-making and Reuse

Rob Sutter, GDIT

Co-authors
Elizabeth Benjamin, Brick Fevold, Craig Palmer and Judith Schofield, GDIT
Louis Blume, EPA

August 29, 2018

Disclaimer: The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.
Desert Tortoise Burrows

- In plot and incidental observations
- Basic data
  - Occupied by tortoise
  - Burrow or den
  - Burrow dimensions
  - Burrow location
  - Burrow condition
- Classify as active or inactive with checklist of evidence
Desert Tortoise, Nevada
Practical Guidance for Integrating Data Management into Long-Term Ecological Monitoring Projects


Robert D. Sutter
Susan B. Wainscott
John R. Boetsch
Craig J. Palmer
David J. Rugg
Application of Quality Assurance and Quality Control Principles to Ecological Restoration Projects

Appendix A: Data Management
Outline of Presentation

• Define data and data management
• Discuss the value of data management
• Introduce elements of data management planning and implementation
Definition of Data

Environmental Data

– Any measurement or information that describe
  • Environmental processes, location or conditions
  • Ecological health effects or consequences
  • Performance of environmental technology

– Includes:
  • Primary data: directly from measurements
  • Secondary data: existing data

(USEPA 2002)
Data Management

• A structured process that promotes data quality, availability, and preservation for analysis, informed decision-making and data reuse.

Three Principles:

– Discoverability
– Accessibility
– Usability
Value of Data Management

Essential for:

- Analysis: data available during the life span of a project
- Informed Decision-making: data available for use in making decisions
- Reuse: data available to resample, reanalyze and other uses
Value of Data Management

Restoration Questions Requiring Long-term Data Collection

- Population dynamics
- Dynamics of ecological processes
- Cumulative effect of stressors
- Effectiveness of management and restoration
- Changing Climate

Big Data
Graded Approach

• Recognizes that not all monitoring projects require the same level of detail in their data management plan

• Ensures the rigor of data management planning is commensurate with:
  – Importance of the work
  – Resources
  – Needs of participating organizations
  – Consequences of potential decisions
Important Questions to Ask for a Graded Approach

• What is the intended use of the data?
  – Human health consideration
  – Protection of federally listed species
  – Quantitative assessment of restoration effectiveness

• What is the level of effort supporting the project?
  – Short term or long term
  – Single organization or an interagency task force
  – Small scale ($) or large scale ($$$)
  – Local or national issue

• What is the potential impact?
  – Ecological impacts
  – Economic impacts
  – Legal defensibility of the data
Data Management Plan
Implementation Model

- Project Description, Administration, & Requirements
- Quality Assurance
- Acquisition & Collection
- Documentation & Metadata Creation
- Data Backup
- Organization, Storage, & Security
- Processing & Analysis
- Preservation & Archiving
- Sharing & Reuse
Data Management Plan
Implementation Model
Planning for Data Management

Three Components

• Project description
• Project administration
• Data management requirements
Data Management Plan
Implementation Model

- Acquisition & Collection
- Quality Assurance
- Documentation & Metadata Creation
- Data Backup
- Organization, Storage, & Security
- Processing & Analysis
- Preservation & Archiving
- Sharing & Reuse
- Project Description, Administration, & Requirements
Data Acquisition and Collection

Projects generate and compile environmental data from numerous sources

- Field observations and measurements
- Field images, audio, voucher specimens, laboratory samples
- Secondary sources: soil surveys, community descriptions, field guides
Data Acquisition and Collection

- SOPs (Standard Operating Procedures) – written instructions that document a routine or sampling activity
- Data collection forms
- Field logistics and data conveyance from field to workspace
- Documenting and resolving unknowns
- Form and sample custody protocols
- Training and debriefing
- Verification and validation
Data Management Plan
Implementation Model
Organization, Storage and Security

Data management must eliminate the potential of data loss throughout the project lifecycle

- Workflow to guide storage of data in original and processed formats
- Version control
- Storage methods
- Access and censorship
Data Management Plan Implementation Model
Data Processing and Analysis

• Procedural Steps for Processing
  – Any manipulation of data to convert it into a useable format for subsequent applications

• Procedural Steps for Analysis
  – Preparing data for analysis, validating statistical assumption
  – Document statistical methods and analysis steps
Data Management Plan
Implementation Model
Data Preservation

• Long-term protection and archiving of interim and final data products
  – Data are complete and certified
  – Managed and secure environment
  – Lossless data format

• Data access, distribution and deposition protocols
Data Management Plan
Implementation Model
Data Sharing and Reuse

Sharing data and data products to ensure that they are readily accessible and can be used internally and by the broader scientific community

• Discoverability
• Accessibility
• Usability
Data Management Plan
Implementation Model
Summary

- Proactive integration of data management into research
- Model provides a guide to the elements of data management
- Quality assurance is part of every element
- Final product of research is not only a manuscript, but the whole data management document
Thank You

Questions