Data Management SOP for the Tampa Bay Estuary Program

Marcus W. Beck

2021-01-07

Contents

Chapter 1

Prerequisites

Chapter 2

Background

2.1 Importance of data

- Data are the foundation of all research products and management decisions
- A data definition, e.g., raw information in flat files, synthesized/derived datasets, models, etc.
- How data are used in applied research/environmental sciences

2.2 Why we need to effectively manage data

- What happens when data are not managed properly - Figure 1, bit-rot
- Professorware
- Benefits of a data management workflow
- Applications in open science

2.3 Goals/objectives of this document

- What it is, what it is not - including what makes TBEP different from other organizations, i.e., we have hands in lots of projects vs one central product (e.g., OHI), so our SOP needs to be generalizable
- Develop pathway for metadata
- Intended audience

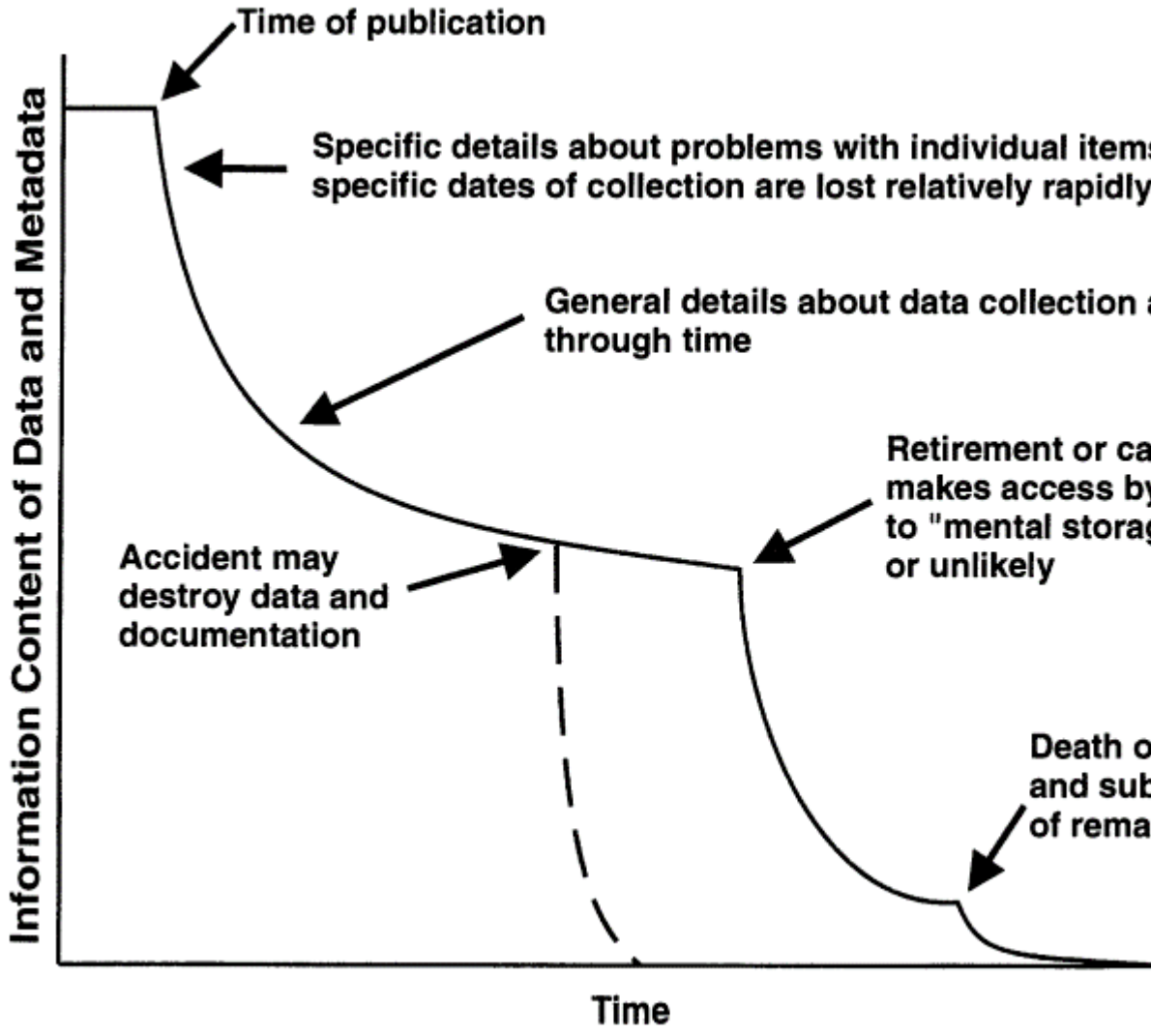

Figure 2.1: Loss of information over time in the absence of data management [@Michener97]

Chapter 3

Key Concepts and Principles

3.1 General concepts

- What are data, i.e., from the perspective of the researcher/agency scientist/manager?
  - Workflow (e..g, operationalizing Twitter scraping), dataset (field/lab data), model products, etc.
  - Ask yourself, who is going to use this and how do I make their lives "easier" by opening the data using FAIR principles?
- The FAIR principles (very broad, emphasize throughout), also general open science definition and how data relates to open science (channel PeerJ paper distinction)

3.2 Specific concepts

- Specific concepts (particularly for tabular data)
  - an overview of tidy data, can a machine read it?
  - normalized tables (including discussion of key variables), what are unique ids (e.g., tberf oyster, how did I make the unique id?), facilitate standard DB queries
  - metadata documentation (min requirements, relevant standards)
- types of data products (e.g., raw data, models, synthesized/derived data, etc.) or types of data (flat file, spatial, disparate)
- how-to cookbooks for data prep (could speak to different parts of the analysis workflow, e.g., project inception, mid-project, post-project/damage control) for archival, naming conventions (e.g., no spaces, short but descriptive, etc.), data dictionaries

- where do data live long-term, what's a doi, considering a data paper, federated repository, etc.
  - GitHub repository
  - Stable URL
  - Official repository

Chapter 4

Data Management Workflow

- Setup some kind of flow chart (if this, then that)
  - What type of project am I working on?
  - What types of products am I expecting?
  - Where am I at with the project (beginning, middle, end/damage control)?
  - How do I want to make the data accessible?

Chapter 5

Case Studies

Demonstrate the workflow

5.1 TBERF oyster restoration project

5.2 DeSoto/RESTORE project

5.3 Red Tide Twitter repo

Chapter 6

Final Words

- emphasis on “something is better than nothing”, fully open is ideal but difficult to achieve
- Just remember FAIR

Chapter 7

Appendice

7.1 Guidelines for best data management practices

7.2 Data types

- Field data
  - Survey forms
  - Tabular*
  - Database of tables
- Database
  - Synthesis
  - New
- Model
  - Actual model
  - Model results
- Dashboard

7.3 Definitions

Dashboard Data Aggregation vs. synthesis Model Tabular Database Flat file Tidy data

7.4 Metadata templates

General - Who, what, when, where, why?

Specific - XML, EML, etc.