Data Management Plans

Bill Michener

University Libraries and Biology Dept.

University of New Mexico

Planning …

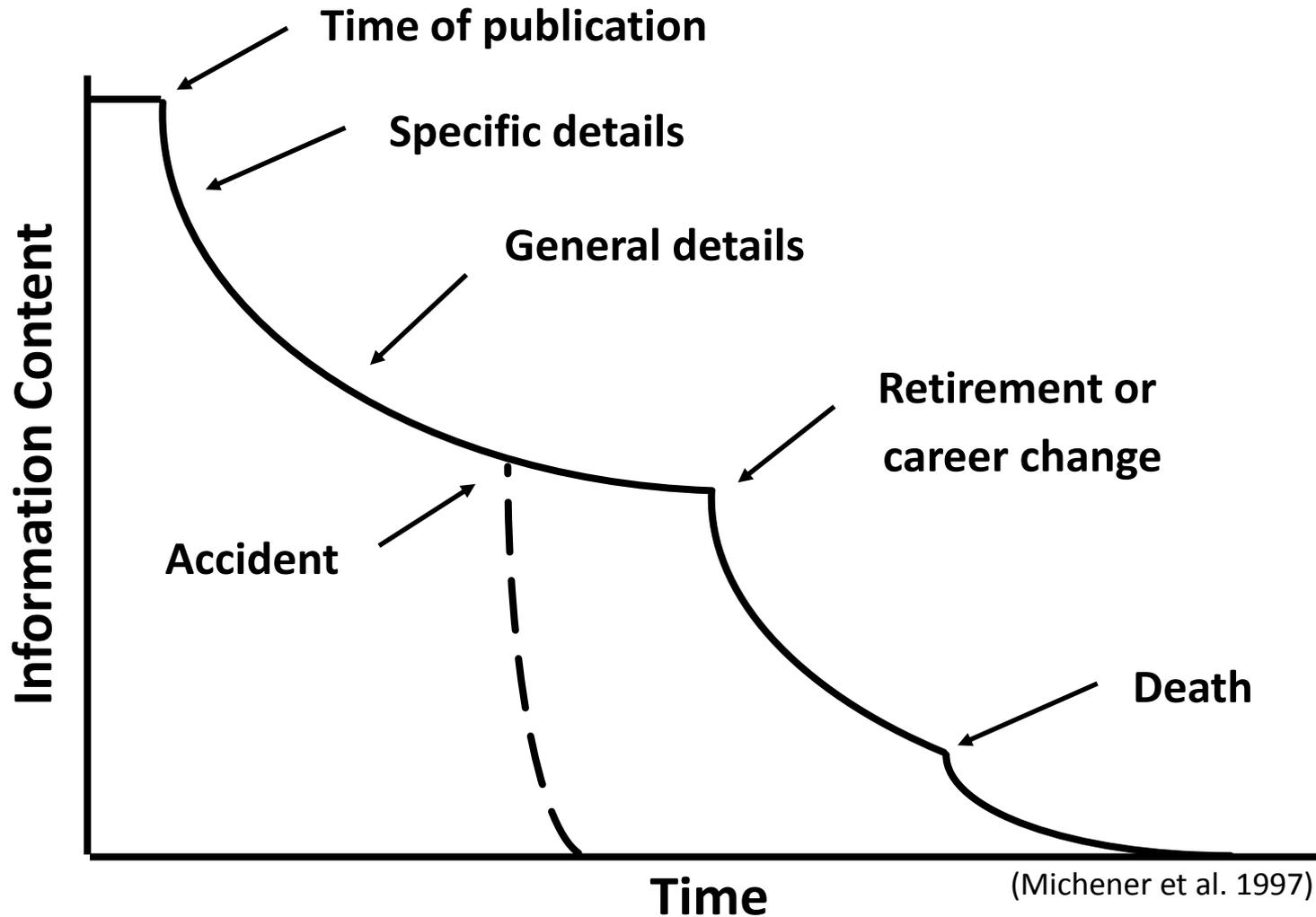"A goal without a plan is just a wish."

Antoine de Saint-Exupery **(1900 - 1944)**

"Plan" to forestall data entropy and promote science

Time of publication

Specific details

General details

Retirement or career change

Information Content

Accident

Death

Time

(Michener et al. 1997)
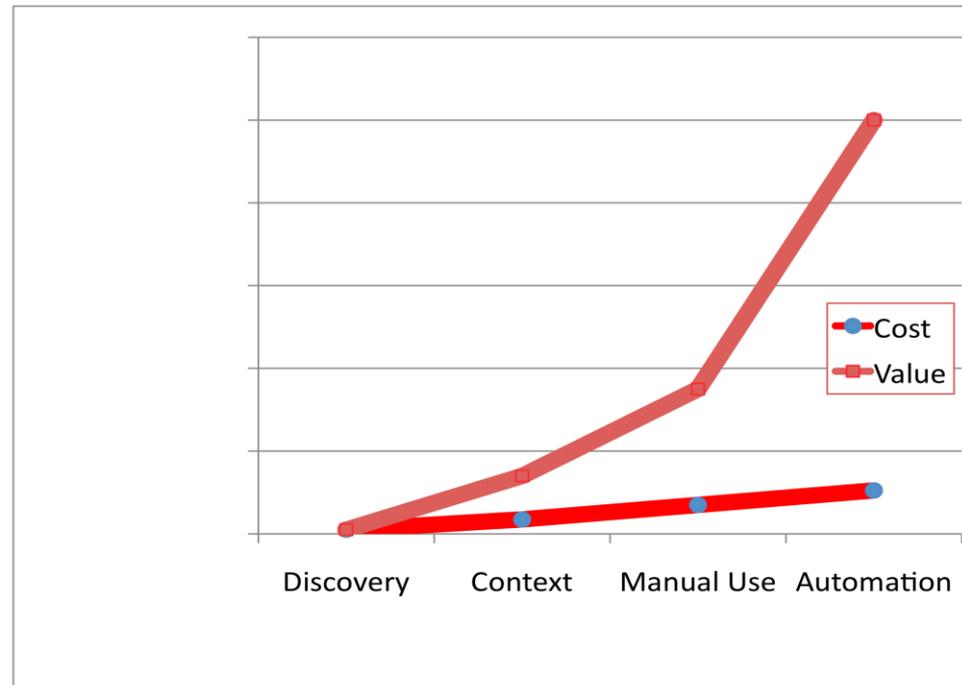
Data as a Product of Science

- Data have value

- A good plan and comprehensive metadata enable data to retain their value

Comparison of Data Management Plan Components

COMPONENT	MIT	DCC	ANDS	CESSDA	JPL	QUT	DAAC
OBJECT CHARACTERISTICS							
Data type (raw, computational, observational)	✓	✓	✗	✗	✗	✓	✓
Naming conventions	✓	✗	✓	✓	✗	✓	✓
Format types	✓	✓	✓	✗	✓	✓	✓
CONTEXTUAL INFORMATION							
Software (resilience, obsolescence)	✓	✓	✗	✗	✗	✓	✓
Required and stored metadata	✓	✓	✓	✗	✗	✓	✓
Documentation (may relate to metadata)	✓	✓	✓	✓	✓	✓	✓
Intellectual control (copyright, ownership)	✓	✓	✓	✓	✗	✓	✓
ENVIRONMENTAL / STORAGE FACTORS							
Data retention	✓	✓	✓	✗	✗	✓	✓
Anticipated volume	✓	✓	✗	✗	✓	✗	✓
Backup	✓	✓	✓	✓	✗	✓	✓
USAGE							
Access restrictions	✓	✓	✓	✗	✗	✓	✗
Privacy / anonymization	✓	✓	✓	✓	✗	✓	✓
Security	✓	✓	✓	✓	✗	✓	✓
Audience	✓	✓	✓	✗	✗	✗	✓

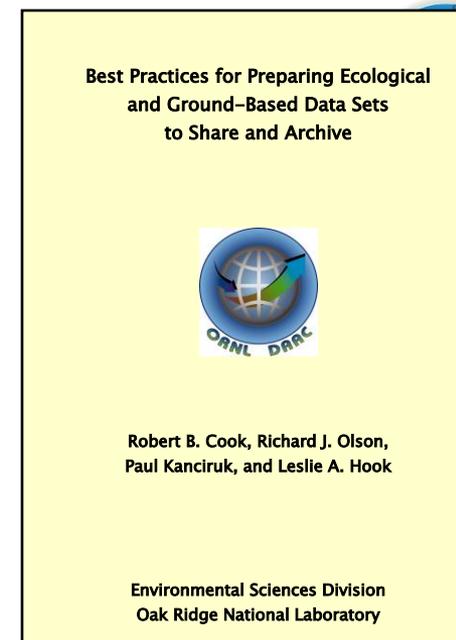Best Practices

— Valerie Enriquez, DataONE Intern

— UC Curation Center

—-- MIT: Massachusetts Institute of Technology -- DCC: Digital Curation Centre -- ANDS: Australian National Data Service

—-- CESSDA : Council of European Soc. Sci. Data Archives -- JPL: Jet Propulsion Laboratory -- QUT: Queensland University of Technology

—-- DAAC: ORNL Distributed Active Archive Center

Recommendations

- Follow best practices
- Start early
- Keep it simple
  - Concise and clear
- Make it a team sport
  - Engage students and colleagues
- Review frequently
  - e.g., quarterly, before every field season, ...
- Minimize or eliminate restrictions on use
- Tailor the plan to the size of the research project

Best Practices

Best Practices for Preparing Ecological and Ground-Based Data Sets to Share and Archive

Robert B. Cook, Richard J. Olson, Paul Kanciruk, and Leslie A. Hook

Environmental Sciences Division
Oak Ridge National Laboratory

Proposed Data Management Plan Template

1. Identification Information

2. Basic Project Information

3. Data Management Policies and Procedures

1. Identification Information

- <u>Title of Plan</u> **–** for example, "Data Management Plan for *such and such a (sub-) project or (sub-)program*"

- <u>Author(s)</u>

- <u>Date</u> – date plan was produced

- <u>Revision History</u> – if appropriate (for example, table that summarizes changes or revisions to plan, date of revision, and author(s) of revisions)

2. Basic Project Information

- <u>Project Name</u> – Name of project associated with individual, group, or consortium

- <u>Key Project Participants and Affiliation</u> – List of project personnel and their organizational affiliations

- <u>Funding Agency(ies)</u> – Organization(s) that support the project

- <u>Project Objectives and Context</u> – Brief summary of project objectives and context

3. Data Management Policies and Procedures

- <u>Guiding Policy(ies) for Project, Institution, and/or Funding Agency</u>:
  - *Data sharing policy* – policies that define what data will be shared, when they will be shared, and any restrictions on use
  - *Data citation policy* – how data should be cited when used by others
  - *Data preservation policy* – what data will be retained, how long will they be retained, and where they will be deposited
  - Other relevant policies
- <u>Data Acquisition, Processing, and Quality Assurance/Quality Control</u> – Procedures and protocols typically employed in acquiring and managing data
- <u>Metadata Content Standards and Metadata Management</u> – Identification of Metadata Content Standards adhered to and approaches for creating and managing metadata
- <u>Data Preservation</u> – Plan that describes what data will be maintained beyond the life of the project, how long they will be preserved, and where and when those data and metadata will be deposited

Example: 1. Identification Information

- Data Management Plan for *Sevilleta LTER Primary Productivity Study, 1997-2003*

- Jane Smith, PI

- 29 December 2003

- <u>Revision History</u>

  - Version 1.0 – 1 June 1997 – Jane Smith

  - Version 1.5 – 15 March 1998 – revised QA/QC – Jane Smith

  - Version 2.0 – 29 December 2003 – included archive/preservation information – Jane Smith & Walker Branch

Example 2. Basic Project Information

- <u>Project Name</u> – Sevilleta LTER Program
- <u>Key Project Participants and Affiliation</u> –
  - Jane Smith (PI) – University of New Mexico (UNM)
  - Walker Branch (MS Student) – UNM
  - Gayle Moore (MS Student) -- UNM
- <u>Funding Agency(ies)</u> – NSF BIO 97-0014
- <u>Project Objectives and Context</u> – "To observe and understand natural changes in primary production under different precipitation …."
  - 1 paragraph to multiple pages (e.g., amplified intro to a paper)

Example 3. Data Management Policies and Procedures

- <u>Guiding Policy(ies) for Project, Institution, and/or Funding Agency</u>:
    - *Data sharing policy* – "Data are freely and openly accessible at the LTER data portal within six months of collection"

Example 3. Data Management Policies and Procedures

- <u>Data Acquisition, Processing, and Quality Assurance/Quality Control</u> – Procedures and protocols typically employed in acquiring and managing data

  - For example, see Cook et al., 2001
  - Point to reference(s)
  - 1 page to many pages (e.g., expanded materials and methods section for data)

Example 3. Data Management Policies and Procedures

- <u>Metadata Content Standards and Metadata Management</u>

    – "Ecological Metadata Language version 1.2"

    – "Metadata entered using *morpho* and stored at _____"

Example 3. Data Management Policies and Procedures

- <u>Data Preservation</u> – Plan that describes what data will be maintained beyond the life of the project, how long they will be preserved, and where and when those data and metadata will be deposited
    - "Data are available through Dryad and DataONE for perpetuity"
    - "Data and metadata were deposited in Dryad in May 2010 in conjunction with publication of ……"

Thanks!