

Scientific Stewardship



What Is It and What Does It Mean To Us?

Ge Peng

CICS-NC/NCEI-NC

(ORCID: 0000-0002-1986-9115; @DrPengAtAVL)

1st CICS Science Conference, 29 December 2016



What Is Scientific Stewardship?

• All activities that ensure and improve the quality and usability of data products

(National Research Council (2007): Data Stewardship encompasses all activities that preserve and improve the information content, accessibility, and usability of data and metadata)

What Is Data Quality?

How good or bad a data product is.

What Is Data Usability?



How easy or hard a data product is understood and used.

Who Should Care?



Who Should Care?

- All Product Key Players everyone who develops, creates, produces, stewards, manages, publishes, or serves the product
- Other major product stakeholders including sponsors, power users, and management
- General users

Why Should We Care?



Good quality is crucial in making sound & effective decisions!

Potential Impacts of Poor Data Quality



(Image source: hediip.ac.uk/wp-content/uploads/Poor-Data-Quality-Pic1.png)

Before:

We didn't usually worry about scientific stewardship

after the data products have been produced.

Now We Need To!





Ensuring and Improving Quality and Utility of Information Is the Law!

- U.S. Law 106-554 (2001) Information Quality Act
- OMB (Office of Management and Budget) Guidelines (2003)
- OSTP (Office of Science and Technology Policy) Memo (2013)



They Are Also Agency's Requirements

- NOAA Administrative Order 212-15 (2008)
- NOAA Plan for Increasing Public Access to Research Results (2015)
- NOAA Data and Publication Sharing **Directive** (2016)

Not Just Scientific Quality Any More

- *Scientific quality*, defined in terms of accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose).
- **Product quality** measures the degree to which the scientific quality is assessed and documented; how accurate, complete and up-to-date the metadata and documentation are.
- **Stewardship quality** assesses the current state of how datasets are documented, preserved, stewarded, and made accessible publicly.
- Use/Service quality measures how easy it is for users to discover, get, understand, trust, and use a given data product and metadata, as well as the state of user service provided by data distributors.

(Source: Ramapriyan et al., 2016, SciDataCon2016)

Stewardship

Ensuring Data Quality and Usability Is an End-to-End Process

Define/Develop/Validate

Create/Evaluate/Obtain

Product

Maintain/Preserve/Access

Use/User Service

Service

Science



Changing Environment

- Data Asset Trustworthiness
- Open Data Transparency
- Data Sharing Fast access and easy to use
- Big Data 4V (Volume, Variety, Velocity, Veracity) systematic and scalable

Data users do not have to wait for or use your data any more!



What Does It Mean To Us?

For principal investigators, data producers, or data providers:

- Documentation, Documentation, Documentation, ...
- Domain Best Practices.

What Does It Mean To Us?

If you serve as scientific stewards: (product SMEs managing data quality and usability)

- Evaluation and Verification,
- Characterization,
- Documentation,
- Data quality metadata, ...

For more details in those and other roles and their high-level responsibilities: http://tinyurl.com/RolesRs-DQU

(Source: Peng et al., 2016, D-Lib Magazine, May/June issue)



It is not just about delivering data any more; it is more about providing information!

-- John Bates, 8 November 2016

Your Practices v.s. Domain Best Practices

Assessment Models

Define/Develop/Validate	Create/Evaluate/Obtain	Maintain/Preserve/Access	Use/User Service
Science	Product	Stewardship	Service
Data System	Product	Stewardship	Service
Maturity Matrix	Maturity Matrix	Maturity Matrix	Maturity Matrix
EUMETSAT	Bates and Privette	Peng et al.	Arndt and Brewer
(2013; 2015)	(2012)	(2015)	(2016)

(More info: Peng et al., 2016, AGU 2016 Fall)



Take Away Messages

- It's an **exciting time** to do what we do!
- We may have to change the way we do things.



- This presentation only scratches the surface.
- We can be passive let others define best practices for us
 OR
- We can be pro-active and participate in the process

THANK YOU

Ge.Peng@noaa.gov



Extra Stuff

Main Requirements for Us

- Be the part of the data stewardship process,
- Be timely in assessing quality and characterization of data,
- Be efficient (systematic and scalable) in ensuring and improving data quality,
- Making data and information about the data readily available and accessible (human-understandable and machine-readable) in a timely fashion.

What Does

Scientific Data Stewardship Mean?

Ensure your data are

- preserved and secure
- available, discoverable, and accessible
- credible and understandable
- usable and useful
- sustainable and extendable
- citable and traceable





DSMM Defines Measureable, Five-Level Progressive Practices

Maturity Scale Key Component	Level 1 - Ad Hoc Not Managed	Level 2 - Minimal Managed Limited	Level 3 - Intermediate Managed Defined, Partially Implemented	Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 5 - Optimal Level 4 + Measured , Controlled , Audit
Preservability	The state of dataset being preservable				
Accessibility	The state of dataset being publicly searchable and accessible				
Usability	The state of data product being easy to understand and use				
Production Sustainability	The state of data production being sustainable and extendable				
Data Quality Assurance	The state of data product quality being assured/screened				
Data Quality Control /Monitoring	The state of data product quality being controlled and monitored				
Data Quality Assessment	The state of data product quality being assessed				
Transparency /Traceability	The state of data product being transparent, trackable, and traceable				
Data Integrity	The state of data integrity being verifiable				

(Data system integrity is also very important but not included in the matrix due to potential security risks to the system.)

How Did We Get here?





DSMM Follows CMMI level Structure



Tiers of Maturity Assessment within Context of Scientific Data Stewardship





Environmental Data Management at NOAA: Archiving, Stewardship, and Access

Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA

National Research Council, 2007

PRINCIPLE #6: Data and metadata require expert stewardship

Data stewardship encompasses all activities that **preserve** and **improve** the information content, **accessibility**, and **usability** of data and metadata.

Non-Functional Requirements on Climate Data Records

(No Particular Order)



Review of NOAA's Plan for
the Scientific Stewardship
Program
Committee on Climate Data

Committee on Climate Data Records from NOAA Operational Satellites

National Research Council, 2005

Long-term CDR stewardship is important

Stewardship of the scientific data

rw of NOAA's Plan fo the Scientific Data

ewardship Program

• Scientific vigilance: Scientists provide oversight of data stewardship

OSTP Memo

Office of Science and Technology Policy

Feb 22, 2013

Subject: Increasing Access to the Results of Federally Funded Scientific Research

Principles: made **available** and **useful** for the public, industry, and the scientific community.

Information Quality Act (Publ. Law 106-554, 2001)

 OMB to provide policy and guideline for ensuring and maximizing the quality, objectivity, utility, and integrity of information

OMB (Office of Management and Budget) Guidelines (2002)

- Data quality, objectivity, and integrity (security of information)
- Transparency high degree for influential information
- Reproducibility

NOAA Administrative Order 212-15 (2008)

Management of environmental and geospatial data and information

- Include end-to-end data stewardship in data management planning
- Take appropriate steps to ensure acceptable accuracy, precision, representativeness, documentation, and long-term continuity of NOAA's quality data sets for the user community

Policy: U.S. Law and Federal Mandates

- U.S. Law 106-554 (2001) Information Quality Act
 - OMB to provide policy and guideline for ensuring and maximizing the quality, objectivity, utility, and integrity of information

• OMB (Office of Management and Budget) Guidelines (2003)

Guideline on data quality, objectivity, integrity, transparency, and reproducibility

OSTP (Office of Science and Technology Policy) (2013)

- Increasing access to the results of federally funded scientific research,
 - Requiring data be made available and useful for the public, industry, and the scientific community

Agency's Requirements: – NOAA and Expert Bodies Guidelines

- National Research Council (2005) Review of NOAA's Plan for the Scientific Stewardship Program
 - Scientific vigilance is critical: Scientists provide oversight of data stewardship
- NOAA Administrative Order 212-15 (2008)
 - Include end-to-end data stewardship in data management planning
 - Take appropriate steps to ensure acceptable accuracy, precision, representativeness, documentation, and long-term continuity of NOAA's quality data sets for the user community
- NOAA Plan for Increasing Public Access to Research Results (2015) and Data and Publication Sharing Directive (2016)
 - Make data produced under a grant, cooperative agreement, and contract publicly accessible in a timely fashion

What Does It Mean To Us?



and quidance on data use

(Source: Peng et al., 2016, D-Lib Magazine, May/June issue)

What Does It Mean To Us?



For other roles and their high-level responsibilities: http://tinyurl.com/RolesRs-DQU

What Is Scientific Stewardship?



(Source: Peng et al., 2016, D-Lib Magazine, May/June issue)

What Is Scientific Stewardship?



(Source: Peng et al., 2016, D-Lib Magazine, May/June issue)