

A Stewardship Maturity Matrix for Environmental Data

Ge Peng¹ and Jeffrey L. Privette²

¹ Cooperative Institute for Climate and Satellites, NC State University (CICS-NC) and NOAA's National Climatic Data Center (NCDC)/Remote Sensing Application Division (RSAD), Asheville, North Carolina, USA

² NOAA/NCDC/Climate Services and Monitoring Division (CSMD), Asheville, North Carolina, USA

Why Do We Need a Maturity Matrix?

U.S. Law (Data Quality Act, 2001) requires and expert bodies recommend that environmental data be:

- Preserved and sustainable
- Secure and accessible
- Transparent and traceable
- Assessed, improved, and scientifically defensible

Currently, there is no systematic framework for assessing the quality of stewardship applied to environmental datasets and providing consistent data integrity and usability information to users and stakeholders.

What Does the Maturity Matrix Do?

Based on published guidance from expert bodies, the stewardship maturity matrix

- Defines key components of scientific data stewardship
- Defines a five-level graduated maturity scale for each component, representing Ad Hoc, Minimal, Intermediate, Advanced, and Optimal stage
- Enables a consistent measure of the stewardship practices applied to diverse products

Contact Us

Ge.Peng@noaa.gov
Jeff.Privette@noaa.gov

What Is Scientific Data Stewardship?

- All activities to preserve and improve the information content, accessibility, and usability of data and metadata (NRC, 2007)
- Activities to ensure and improve the quality and usability of environmental data

Reference

National Research Council (NRC), 2007: Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 130 pp.

Who Could Use the Matrix?

- Environmental data providers and stewards seeking to evaluate and improve the quality and usability of their products
- Modelers, decision-support system users, and scientists seeking to better understand the upstream data and data quality management practices applied to their input datasets

Maturity Matrix for Long-Term Scientific Data Stewardship (Draft)

Maturity Scale	Preservation	Accessibility	Usability	Production Sustainability	Data Quality Assessment	Data Quality Monitoring	Transparency and Traceability	Information Integrity
Level 1 - Ad Hoc Not Managed	Any storage location Data only	Not publicly available Person-to-person	Extensive product-specific knowledge required No document online	Ad Hoc No obligation or deliverable requirement	Algorithm theoretical basis assessed	Sampling and analysis are spotty and random in time and space	Limited product information available Person-to-person	Unknown data integrity Unknown IT system
Level 2 - Minimal Managed Limited	Non-designated repository Redundancy Limited metadata	Publicly available Not searchable online	Non-standard data format Limited document (e.g., user's guide) online	Short-term Individual PI's commitment (grant obligations)	Level 1 + Research product assessed	Sampling & analysis are regular in time and space Limited product-specific metric defined & implemented	Product information available in literature	No data integrity check Non-designated IT system
Level 3 - Intermediate Managed Defined & Partially Implemented	Designated repository/archive Conforming to community archiving process and metadata Conforming to limited archiving standards	Available online Limited data server performance Collection/dataset searchable	Community standard-based interoperable format and metadata Documentations (e.g., source code, product algorithm document, data flow diagram) online	Medium-term Institutional commitment (contractual deliverables with specs and schedule defined)	Level 2 + Operational product assessed	Regular/ frequent, and systematic Not automatic Community metric defined and partially implemented	Algorithm Theoretical Basis Document (ATBD) and source code online Dataset configuration managed (CM) Data citation tracked (e.g., utilizing DOI system) Unique object identifier (UOI) assigned (dataset, documentation, source code)	Data ingest & archive integrity verifiable (e.g., checksum technology) Designated IT System Conforming to low IT security requirements
Level 4 - Advanced Managed Well-Defined & Fully Implemented	Level 3 + Conforming to community archiving standards	Level 3 + Enhanced data server performance Granule/file searchable	Level 3 + Basic capability (e.g., subsetting, aggregating) and data characterization (overall/global, e.g., climatology, error estimates) available online	Long-term Institutional commitment Product improvement process in place	Level 3 + Quality metadata assessed Limited quality metadata	Level 3 + Anomaly detection procedure well-documented and fully implemented using community metric, automatic, tracked, and reported	Level 3 + OAD (Operational Algorithm Description) online (CM + UOI)	Level 3 + Data access integrity verifiable (e.g., checksum technology) Conforming to moderate IT security requirement
Level 5 - Optimal Level 4 + Measured, Controlled, and Audit	Level 4 + Archiving process performance controlled, measured, and audited Future archiving process and standard changes planned	Level 4 + Dissemination reports available Future technology changes planned	Level 4 + Enhanced online capability (e.g., visualization, multiple data formats) Community metric set of data characterization (regional/cell) online External ranking	Level 4 + National or international commitment Changes for technology planned	Level 4 + Assessment performed on a recurring basis Community quality metadata External ranking	Level 4 + Cross-validation of temporal & spatial characteristics Consistency check Dynamic providers/users feedback in place	Level 4 + System information online Complete data provenance	Level 4 + Data authenticity verifiable (e.g., data signature technology) Conforming to high IT security requirement Performance of data and system integrity check monitored & reported Future technology change planned

Many subject matter experts of archive, access, user service, system engineering and architecture, software engineering, IT security, data management, configuration management, satellite data product development, and research-to-operation transition at or affiliated with NOAA's NCDC have provided their knowledge in the related fields. Many of them have also reviewed and provided feedback on the definitions of maturity levels of one or more key components. We would like to thank them all, especially Steve Ansari, Drew Saunders, John Keck, Richard Kauffold, Bryant Cramer, Jason Cooper, Ken Knapp, Nancy Ritchey, Ed Kearns, M. Scott Koger, John Bates, and Otis Brown.

NOAA's National Climatic Data Center, Asheville, North Carolina

Protecting the past... Revealing the future

