The State of Building a Consistent Framework for Curation and Presentation of Earth Science Data Quality

Ge Peng[1], Hampapuram Ramapriyan[2], and David F. Moroni[3]

[1] Cooperative Institute for Climate and Satellites–North Carolina (CICS-NC) and NOAA National Centers for Environmental Information (NCEI), Asheville, NC

[2] Science Systems and Applications, Inc. (SSAI) and NASA Goddard Space Flight Center (GSFC), Greenbelt, MD; [3] Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA

Background

Improving collection, description, discovery, and usability of information about data quality in Earth science data products is critical in ensuring data use but requires coordinated efforts of people from multiple disciplines.

Under the auspices of the Federation of Earth Science Information Partners (ESIP), the Information Quality Cluster (IQC) brings together national and international data quality management researchers and practitioners from various disciplines to evaluate and establish best practices and standards for data quality for the Earth science community. The IQC evaluates community data quality best practices and standards; makes recommendations for improvement in various aspects of managing data quality in Earth science data products; ensures that producers of data products are aware of standards and best practices for conveying data quality, and data providers/distributors/intermediaries establish, improve, and evolve mechanisms to assist users in discovering and understanding data quality information; and provides guidance to data managers and stewards on how best to implement data quality standards and best practices to ensure and improve maturity of their data products.

This presentation will provide an overview of the current state of building a consistent framework for curating and presenting Earth science data quality in terms of science, product, stewardship, and service maturity of individual datasets.

Outline of Agency and International Community Activities

- **ESIP Activities**
  - Information Quality Cluster – focused on identifying challenges, use cases, representation of DQ (Data Quality)/IQ (Information Quality) to help users
- **Other relevant activities**
  - NASA Earth Science Data System Working Groups (ESDSWG)–Metrics Planning and Reporting WG (Product Quality Checklists)–2010–2012
  - NASA ESDSWG Data Quality WG (Recommendations)–2014–present
  - NOAA Algorithm, Product, Stewardship, and Service Maturity Matrices–2008–present
  - EUMETSAT CORE-CLIMAX System Maturity Matrix and GAIA-CLIM Measurement Maturity Matrix
  - CEOS Essential Climate Variables (ECV) Inventory Questions
  - GEOSS Data Quality Guidelines
  - Quality Assurance framework for Earth Observation (QA4EO)
  - ISO Metadata Quality Standards (19157:2013; 19158:2012)
  - NCAR Community Contribution Portal

Types of Earth Science Data Quality

- ***Scientific quality*** is defined in terms of accuracy, precision, uncertainty, validity, and suitability for use (fitness for purpose);
- ***Product quality*** takes the following considerations into account: 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*** addresses questions such as how well data are being managed, preserved, accessed, and cared for;
- ***Service quality*** deals with how easy it is for users to find, get, understand, trust, and use a given data product along with its metadata, as well as ensuring an archive has the requisite knowledge base and people functioning as subject matter experts available to help its data users.

(Based on Ramapriyan et al. 2016)

Dataset Lifecycle-Stages-Based Maturity Assessment Frameworks

Define/Develop/Validate (Science) → **Create/Evaluate/Obtain** (Product) → **Maintain/Preserve/Access** (Stewardship) → **Use/User Service** (Service)

Science Maturity Matrix

EUMETSAT (2013; 2015a); Zhou et al. (2016)

- Developed for assessing the capability of measurement and production systems and algorithm maturity.
- Applied to 37 EU data records of essential climate variables (EUMETSAT, 2015b)
- Applied to 68 S-NPP/JPSS data products (Zhou et al., 2016)

Data/Product Maturity Matrix

Bates and Privette (2012)

- Developed for assessing the completeness of satellite climate data record (CDR) datasets
- Applied to 32 NOAA CDRs (Bates et al., 2015)

Stewardship Maturity Matrix

Peng et al. (2015)

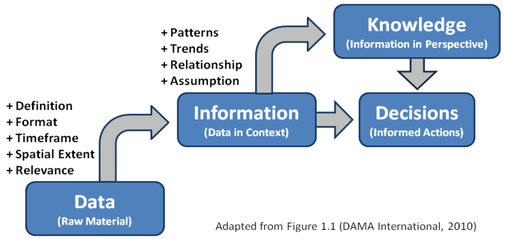
- Developed for assessing maturity of stewardship practices of environmental datasets
- Applied to over 750 NOAA Earth Science datasets (e.g., Peng et al., 2016)

Service Maturity Matrix

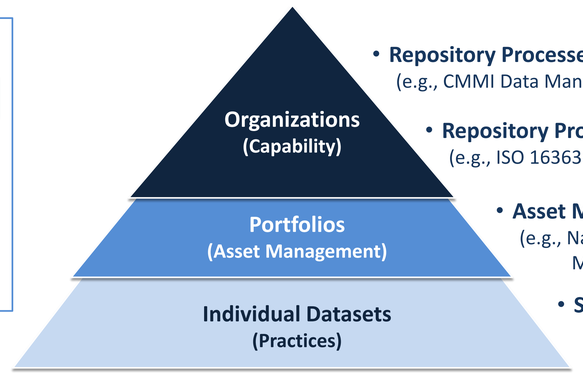
Arndt and Brewer (2016)

- Developed for assessing use and service maturity of environmental datasets
- Under-development by the NOAA/NCEI Service Maturity Matrix Working Group

Pathway to Sound Decisions from Raw Data



Tiers of Maturity Assessment Models



- **Repository Processes Maturity** (e.g., CMMI Data Management Maturity)
- **Repository Procedures Maturity** (e.g., ISO 16363:2012–trustworthiness)
- **Asset Management Maturity** (e.g., National Geospatial Dataset Asset Lifecycle Maturity Model (FGDC, 2016))
- **Stewardship Practices Maturity** (e.g., NCEI/CICS-NC Data Stewardship Maturity Matrix (Peng et al., 2015))

Getting Involved with ESIP Information Quality Cluster

- Information Quality Cluster (IQC) of the Federation of Earth Science Information Partners (ESIP) coordinates and facilitates evaluation of best practices and standards for data quality from the Earth Science community.
- Its vision is to become internationally recognized as an authoritative and responsive information resource for guiding the implementation of data quality standards and best practices of the science data systems, datasets, and data/metadata dissemination services.
- Participation is encouraged and anyone can join by subscribing to the ESIP IQC email list at http://wiki.esipfed.org/index.php/Information_Quality

References

Arndt, D. and M. Brewer, 2016: Assessing service maturity through end user engagement and climate monitoring. *ESIP 2016 summer meeting*. July 19–23, 2016. Durham, NC, USA.

Bates, J. J. and J.L. Privette, 2012: A maturity model for assessing the completeness of climate data records. *EOS*, Transactions of the AGU, 44, 441. doi: 10.1029/2012EO440006.

Bates, J.J., J.L. Privette, E.K. Kearns, W.J. Glance, and X. Zhao, 2015: Sustained Production of Multidecadal Climate Records: Lessons from the NOAA Climate Data Record Program. *BAMS*, 97. doi: http://dx.doi.org/10.1175/BAMS-D-15-00015.1.

DAMA International, 2010: Guide to the Data Management Body of Knowledge (DAMA-DMBOK). Eds. Mosley, M., Brackett, M., & Earley, S., Technics Publications, LLC, New Jersey, USA. 2nd Print Edition.

EUMETSAT, 2013: CORE-CLIMAX Climate Data Record Assessment Instruction Manual. Version 2, 25 November 2013.

EUMETSAT, 2015a: GAIA-CLIM Measurement Maturity Matrix Guidance: Gap Analysis for Integrated Atmospheric ECV Climate Monitoring: Report on system of systems approach adopted and rationale. Version: 27 Nov 2015.

EUMETSAT, 2015b: CORE-CLIMAX European ECV CDR Capacity Assessment Report. Version: v1, 26 July 2015.

FGDC, 2016: National Geospatial Data Asset (NGDA) Lifecycle Maturity Assessment (LMA) 2015 Report - Analysis and Recommendations. 93 pp. Version: 8 December 2016.

Peng, G., J.L. Privette, E.J. Kearns, N.A. Ritchey, and S. Ansari, 2015: A unified framework for measuring stewardship practices applied to digital environmental datasets. *Data Science Journal*, 13. doi:http://dx.doi.org/10.2481/dsj.14-049.

Peng, G., J. Lawrimore, V. Toner, C. Lief , R. Baldwin, N. Ritchey, D. Brinegar, and S. A. Delgreco, 2016: Assessing Stewardship Maturity of the Global Historical Climatology Network-Monthly (GHCN-M) Dataset: Use Case Study and Lessons Learned. *D.-Lib Magazine*. 22, doi:10.1045/november2016-peng

Ramapriyan, H., G. Peng, D. Moroni, and C.-L. Shie, 2016: Ensuring and Improving Information Quality for Earth Science Data and Products – Role of the ESIP Information Quality Cluster. SciDataCon 2016, 11 – 13 September 2016, Denver, CO, USA.

Zhou, L. H., M. **Divakarla, and X. P. Liu,** 2016: An Overview of the Joint Polar Satellite System (JPSS) Science Data Product Calibration and Validation. *Remote Sensing*, 8(2). doi:10.3390/rs8020139

National Oceanic and Atmospheric Administration | National Aeronautics and Space Administration

These activities were carried out across multiple United States government-funded institutions (noted above) under contracts with the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). Government sponsorship acknowledged.