A Platform to Provide International and Inter-Agency Support for Data and Information Quality Solutions and Best Practices

David F. Moroni[1] (david.f.moroni@jpl.nasa.gov); H. K. (Rama) Ramapriyan[2] (Rama.Ramapriyan@nasa.gov); Ge Peng[3]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, [2]Science Systems and Applications, Inc., Earth Science Data Information System Project, Goddard Space Flight Center, Greenbelt, MD,
[3]Cooperative Institute for Climate and Satellites - North Carolina (CICS-NC), NC State University and NOAA National Centers for Environmental Information (NCEI), Asheville, NC, USA

Overview:

The NASA Data Quality Working Group (DQWG) was initiated in March 2014 and has continued through March 2017 resulting in a completion of 3 years of activities. While the efforts within this working group have been substantial from NASA's perspective, others outside of NASA and in the international arena have expressed desire to participate in related activities. This helped to foster the reactivation of the Earth Science Information Partners (ESIP) Information Quality Cluster (IQC) in 2014, which continues to this day. The ESIP IQC has a broader scope compared to the DQWG, both in terms of context and the ability to provide open membership, which has resulted in a growing collaboration from inter-agency and international participants. During this time, the IQC has evaluated new use cases, developed a technical manuscript summarizing its activities and plans for future work, and has facilitated the review of use cases and recommendations from the DQWG and other groups. The IQC has formally introduced definitions of four aspects of information and data quality: scientific, product, stewardship and service. The IQC has defined high-level roles and responsibilities of major players including data producers for ensuring and improving data quality and usability. The IQC continues to advocate for use case submission and evaluation as an effective way to capture and better understand the needs, challenges and capabilities of the Earth science data community. Beginning in 2017 and moving into 2018, the DQWG and IQC are working together on a number of activities including: identifying emerging technologies/practices/solutions for information quality, engaging inter-agency and international communities, more direct feedback from Earth observation missions, extracting additional recommendations from new use cases beyond the NASA perspective, and publishing our findings and recommendations in white papers, conferences, and peer-reviewed journals.

Figure 1: DQWG Historical Legacy of Milestones

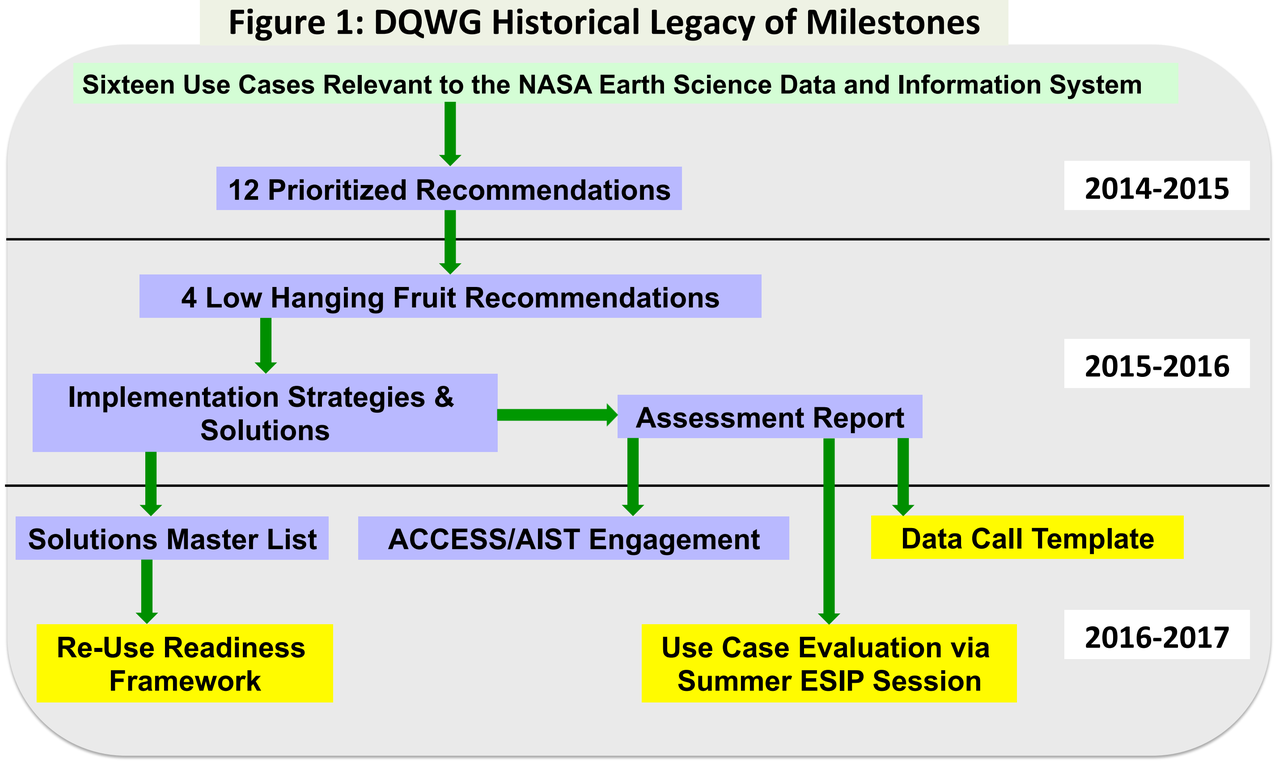


Figure 2: DQ Interest Survey Primary Interest Area for 2017-2018

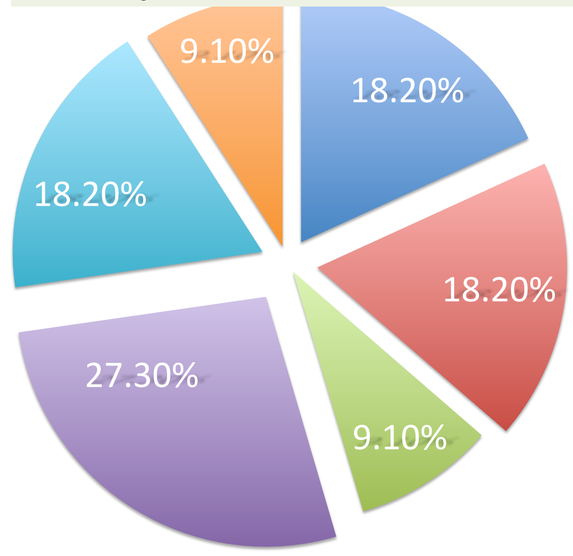


Figure 3: DQWG Action Plan for 2017-2018

Mission Statement
Discover and assess data quality recommendations and solutions in the ***inter-agency*** and ***international*** arena to improve upon existing ***technologies***, practices, and standards in support of ***end-to-end data lifecycle*** stewardship in the NASA Earth science domain.

Stakeholders
- NASA HQ
- ESDIS
- DAACs
- SIPSs
- ACCESS
- AIST
- MEaSUREs
- NASA Earth science instrument teams
- ESIP Information Quality Cluster (IQC)

Approach
- Conduct volunteer pilot study for "Data Call Template".
- Continued volunteer use case contribution and evaluation under leadership of the ESIP IQC.
- Identify new development concepts (e.g., AIST and ACCESS projects) that can be leveraged to facilitate data quality recommendations.
- Use the "Re-use Readiness Framework" to assess the DQWG-endorsed solutions provided in the Solutions Master List.
- Publish previous work in ESDIS-approved outlets.

Outcomes, Deliverables, Milestones
- Operational readiness assessment of "Data Call Template".
- An updated "Solutions Master List".
- Data Quality Section Template for DMP.
- Publish one or more of the highest priority, and most actionable of the previous recommendations/ solutions as an ESO document.
- Publish consolidated recommendations from DQWG annual reports for persistent and public domain access and citation.

Figure 4: ESIP IQC Use Case Evaluation Summary from 2016.

Annotation: 20 use cases evaluated (including 16 legacy DQWG use cases) across DQ Management Phases (columns) and IQ aspects (rows). Some use cases overlap.

	Capture	Describe	Facilitate Discovery	Enable Use
Science	9	16	9	5
Product	11	18	10	5
Stewardship	7	11	6	6
Service	5	10	6	5

Figure 5: IQC Activities for 2015-2016

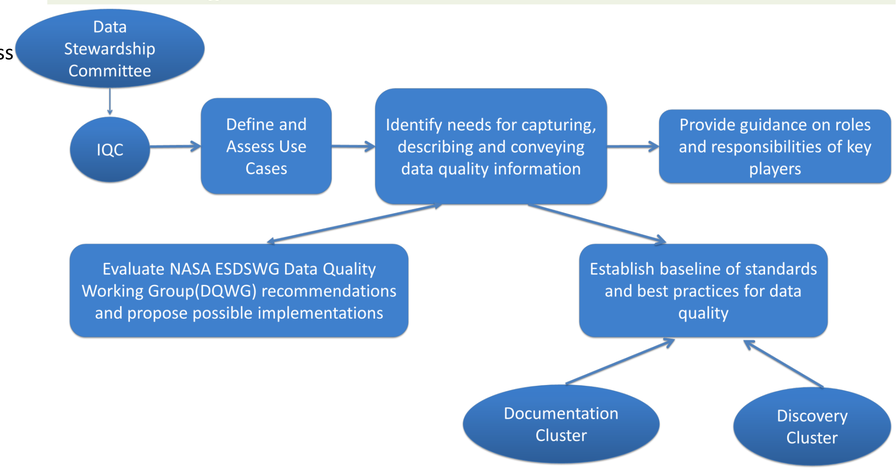


Figure 6: Scope of Mutual Influence and Domain Knowledge

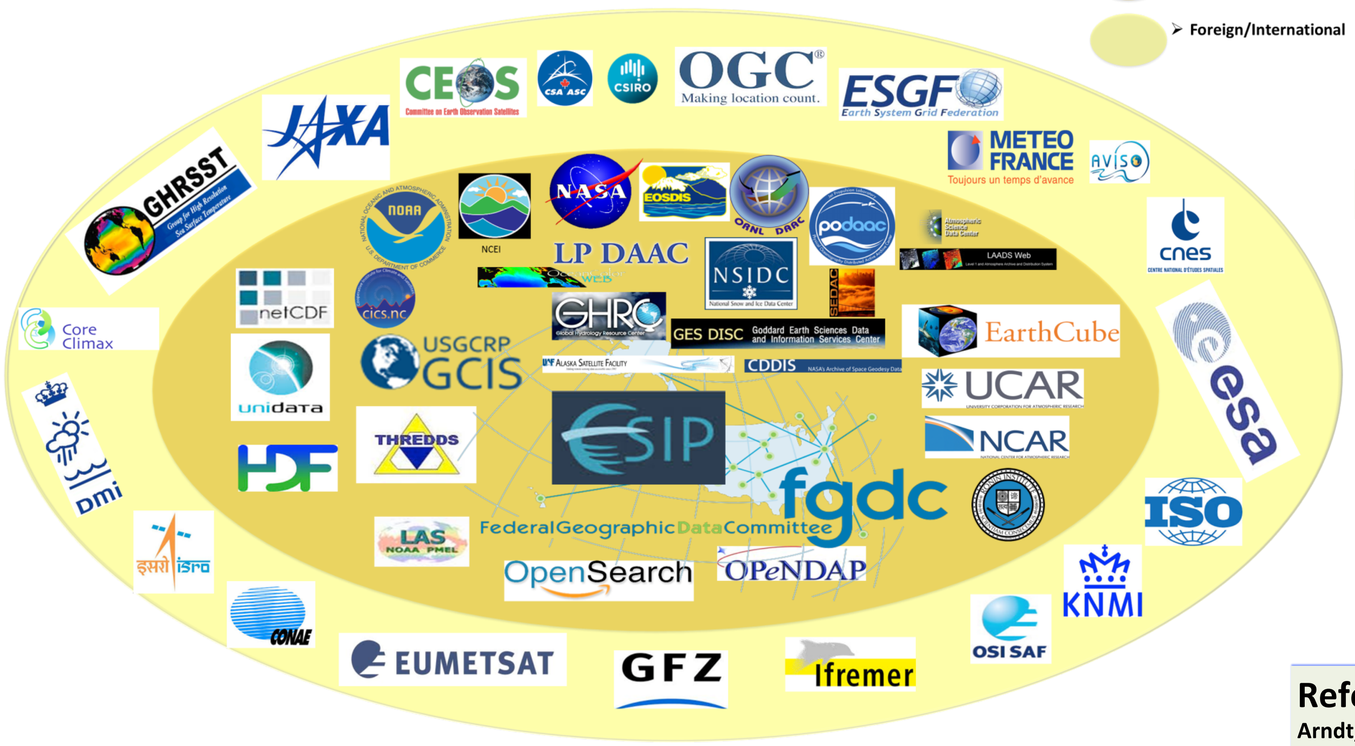
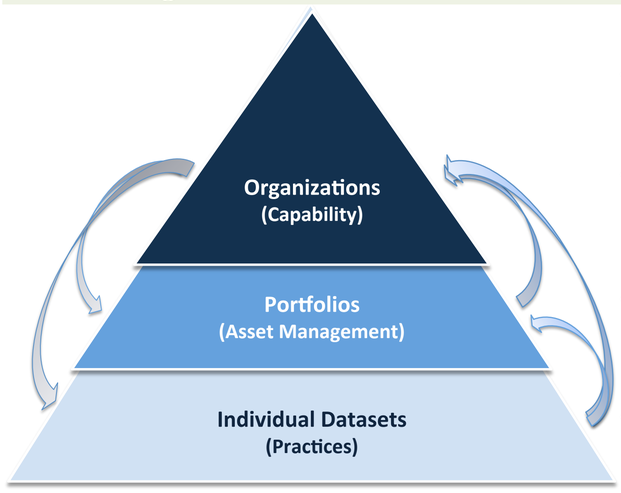


Figure 7: Dataset Lifecycle-Stages-Based Maturity Assessment Models

Define/Develop/Validate Science	Produce/Evaluate/Obtain Product	Maintain/Preserve/Access Stewardship	Use/User Service Service
Science Maturity Matrix	**Data/Product** Maturity Matrix	**Stewardship** Maturity Matrix	**Service** Maturity Matrix
EUMETSAT (2013; 2015a) • Developed for assessing the capability of measurement and production systems for climate data records of essential climate variables • Applied to 37 EU data records of essential climate variables (EUMETSAT 2015b)	**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)	**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)	**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

Figure 8: Tiers of Scientific Data Stewardship Maturity



- **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 Maturity Model)
- **Stewardship Practices Maturity** (e.g., NCEI/CICS-NC Data Stewardship Maturity Matrix)

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. DOI: http://dx.doi.org/10.1175/BAMS-D-15-00015.1

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.

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. 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

Presented at International Ocean Vector Winds Science Team Meeting on May 2-4, 2017 in San Diego, CA. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology. 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.