High-Priority Data Quality Recommendations for Data Producers and Distributors

Status of this Memo

This document provides a set of high-priority recommendations by the NASA Earth Science Data System Working Groups' (ESDSWG) Data Quality Working Group (DQWG) to data producers and distributors of the NASA Earth Science community. The distribution of this document is unlimited.

Change Explanation

Not Applicable.

Copyright Notice

Abstract

This document provides a set of actionable and high-priority recommendations extracted from two years (2014 to 2016) of use case development and evaluation resulting in numerous solutions and recommendations on behalf of members of the DQWG. These recommendations are in the context of data quality and intended for NASA-funded producers and distributors of Earth science data. The focus of this document is on collecting and conveying data quality information to end users rather than providing a precise definition of data quality. It is assumed that data producers and distributors know the meaning of data quality and how it applies to the Earth science data products they produce and/or manage. While originally intended for NASA stakeholders, informal external review of these recommendations at past science and informatics conferences suggests that these recommendations, at least in part, may also be applicable to Earth science data beyond the NASA realm. More robust assessment of these recommendations in their applicability to stakeholders external to NASA will be made possible through this published document.

In the context of this document, the term “data producer” refers to those responsible for creation/production of data that is delivered to a DAAC for archival and distribution. Examples include Science Investigator-led Processing Systems (SIPSs) and NASA mission directed Science Data Systems (SDS) that work closely with the algorithm developers and cal/val teams. Data distributors receive the data from the data producers and make them available to the user community. This document highlights a subset of high-priority recommendations for the Earth Science Data and Information System (ESDIS) Project[1] to plan and coordinate concrete actions

[1] The Earth Science Data and Information System (ESDIS) Project is responsible for management, development and operation of the Earth Observing System Data and Information System (EOSDIS).

to be taken by data producers and distributors (i.e., DAACs) based on the "Prioritized Recommended Implementation Actions" introduced in this document. It points to existing potential solutions that can be adopted across the Earth Observing System Data and Information System (EOSDIS) and the NASA Earth science research community. Please note that these solutions may not be applicable to all types of datasets. For some, including in-situ measurements, concrete implementation solutions still need to be identified/developed.

Table of Contents

---

1. Introduction

The purpose of this document is to describe the high-priority and actionable recommendations, a.k.a. Prioritized Recommended Implementation Actions (PRIAs), that were made by the Data Quality Working Group (DQWG), one of NASA's Earth Science Data System Working Groups (ESDSWG) during 2014-2016, as a result of analyzing a number of data quality use cases related to NASA's Earth science data lifecycle. An accompanying document [10] covers these use cases and the comprehensive set of recommendations from the DQWG in more detail.

The DQWG was formed at the annual meeting of the ESDSWG in 2014 as a result of interest expressed by representatives within the Earth Science Data Information Systems (ESDIS) Project, a subset of managers of the EOSDIS Distributed Active Archive Centers (DAACs), and interested investigators from the Making Earth System Data records for Use in Research Environments (MEaSUREs) Program. The core objectives of the DQWG were to evaluate current processes and make recommendations to the ESDIS Project and the Earth Science Data Systems (ESDS) Program for improvements in the areas of capturing, representing and enabling the use of data quality information describing accuracy, precision, uncertainty and applicability ("fitness for use") such that distinguishability among apparently similar datasets (e.g. ability to distinguish between measurements of the same parameter captured during the same time window in the same area but with different approaches) as well as usability are clear and well-supported. Since its formation, the DQWG has:

- Developed 16 use cases and analyzed them to identify data quality related issues that affect users of NASA Earth science data;
- Arrived at approximately 100 recommendations to address the issues;
- Identified 12 high-priority recommendations;
- Narrowed these down to 4 "low-hanging fruit" (LHF) recommendations;
- Created a "Solutions Master List" (SML) [11] that can address the LHF recommendations;
- Mapped the solutions into generalized implementation strategies;
- Arrived at a set of 8 PRIAs that address the 4 LHF recommendations.

Figure 1 depicts the process used by the DQWG, showing the steps described above. The DQWG report [10] provides a more detailed discussion of this process, along with the full set of recommendations and PRIAs. While the 12 "Prioritized Recommendations" indicated in figure 1 were deemed to be of high-priority, the 4 LHF recommendations were considered to be both highest priority and most "actionable" compared to the remaining 8 recommendations due to the existence of potential solutions that were at a level of operational maturity and portability that established a minimal level of additional effort (i.e., lowest cost) and maximum level of sustained technical support for near-term, operational deployment and implementation at the DAACs. The focus placed on just those 4 LHF recommendations allowed the DQWG to identify appropriate solutions which resulted in the derivation of 8 specific PRIAs. While these PRIAs are the main focus of this document as discussed below, it is worth pointing out for future consideration the other 8 recommendations from the set of 12 prioritized recommendations mentioned above that were not deemed LHF at the time of their evaluation. Thus, it is conceivable that additional PRIAs could be developed with a follow-on investigation; however, this action is beyond the scope of this document.

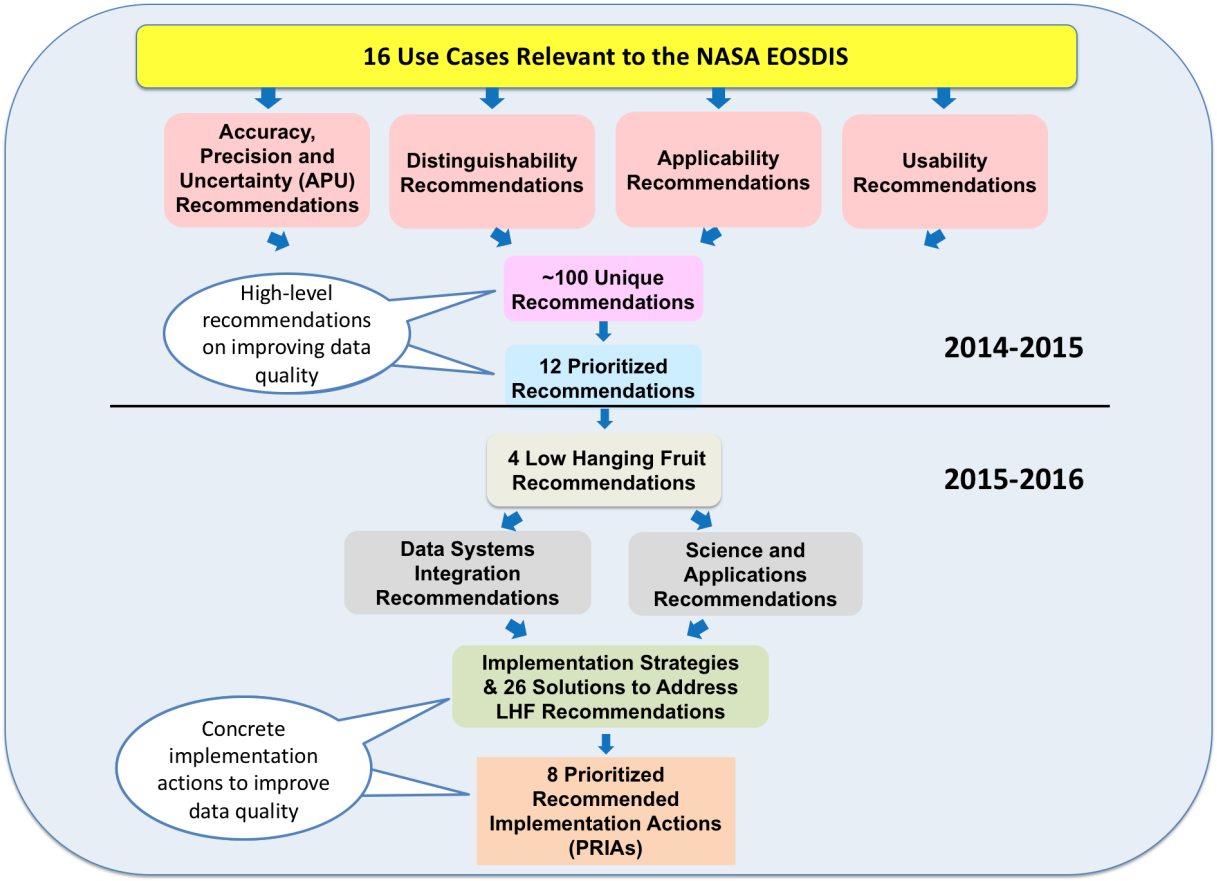

Figure 1. DQWG Historical Legacy of Milestones [10]. Years shown on the right side of the flowchart indicate the time period in which these deliverables were produced. Different colors are used to distinguish between outputs from different stages in the DQWG process.

2. Implementation Strategies

Prior to developing the 8 PRIAs, a Solutions Master List (SML) [11] was built to identify and document solutions to address the 4 LHF recommendations for capturing, describing, enabling use, and facilitating discovery of data quality information and related artifacts. The 26 solutions collected in the SML have been mapped into six implementation strategies. Implementation strategies categorize implementation solutions and provide high-level guidance on approaches to improve the capturing, describing, discovery, and usage of data quality information. Though these strategies are based on the solutions captured for the 4 LHF recommendations, they were purposely generalized to be applicable to a broader set of data quality recommendations. The following subsections provide a brief summary of each of the implementation strategies, which are meant to be broad and high-level. Section 3, Prioritized Recommended Implementation Actions, provides more details on implementation recommendations for data quality.

2.1 Facilitate DAAC-PI Communication
Effective and close communication between DAACs and science teams/PIs is important to ensure effective exchange of thoughts and consolidation of ideas on describing, representing, and using quality information for data to be created by projects and to be archived at the DAACs. This strategy suggests appropriate mechanisms to be developed and leveraged to facilitate such communication throughout the data lifecycle.

2.2 Support Metadata Creation
This strategy points out the need for tools that can support both data providers and data curators at data archives to easily create and/or transform metadata documents to conform to multiple common metadata standards.

2.3 Support Metadata Validation
After data archives receive data/metadata documents from data providers or after data users access data/metadata from data archives, there is a need to validate the completeness of metadata at different levels (e.g., collection and granule) as well as its conformance to multiple common metadata standards. From this validation an investigator may generate reports and/or scores, which can help data archives to identify missing components and also allow data users to better assess both the disparate and common qualities of multiple data products sharing common characteristics.

2.4 Guide, Instruct and Disseminate
ESDIS and DAACs should identify and adopt efficient and consistent ways to help data users access and understand data quality information (e.g., error sources, dataset limitations, and quality assessment) as these would address user questions about data quality and make user feedback about data quality available to user communities and science teams/PIs, if needed.

2.5 Provide User Support Services
User support is important not only for data users to get direct help from experts on accessing, understanding, and using data products, but also for data archives and producers to collect feedback from data users and identify issues of data products based on real data user experience. Such feedback can be further shared with broader user communities and help the usage of data products.

2.6 Consolidate Quality Information Representation
Given the fact that different data products are distributed in different data archives, users usually need data products from more than one archive. This strategy points out the importance of an efficient way to present and convey quality information to data users consistently across the archives.

3. Prioritized Recommended Implementation Actions (PRIAs):

This section describes the 8 PRIAs in detail. Table 1 lists the titles of the 8 PRIAs and their mappings to implementation strategies and solutions. Though these strategies are based on the solutions captured for the 4 LHF recommendations, they are general enough and expected to be applicable for broader data quality recommendations, even beyond the scope of past and current DQWG efforts. Each PRIA is also linked to multiple solutions (referenced

through the index number in the SML). More information about these solutions can be found in the SML summary (Appendix B), with full details available at [11]. The main purpose of this linkage is to provide concrete guidance and existing/implementable solutions for data archives and/or science PIs, who would consider adoption of the recommended implementation actions. Following Table 1, the 8 PRIAs are described in sections 3.1 through 3.8.

Table 1. A synopsis of 8 PRIAs, their associated solution numbers, implementation strategies, and actionees.

PRIA	Solution Numbers (see hyperlink for detail)	Implementation Strategies	Actionees
NASA Recommended Use of ISO Standard	3, 20, 21	Support Metadata Creation	ESDIS
Help users to access and understand data quality information	11, 14, 17, 24	User Support	ESDIS/DAACs
Metadata authoring and validating tools	1, 2, 3, 4, 19, 22, 23	Support Metadata Creation; Support Metadata Validation	ESDIS/DAACs
Develop tools to help users to leverage data quality information	6, 12, 14, 15, 25	User Support	ESDIS/DAACs
Finer-levels of metadata	6, 22, 25, 26	Consolidate Quality Information Representation; Support Metadata Creation	ESDIS
Science Advice	7, 8, 18	Guidance, Instruction, and Dissemination	ESDIS
Facilitate DAAC-PI communication	1, 4, 7, 9, 10	Facilitate DAAC-PI Communication	ESDIS
Data quality best practices	6, 9, 10, 13	Guidance, Instruction, and Dissemination	DAACs

3.1. NASA Recommended Use of ISO Standard

ESDIS should provide appropriate documentation and guidance on how to employ quality attributes of the NASA implementation of the ISO 19157:2013 [7] and 19157-2:2016 [8] standards, once these are fully established. ISO 19115-2:2009 [6] is already broadly implemented across NASA ESDIS, and while we are mindful of this from a metadata standards and completeness perspective, we place more emphasis on the quality-specific 19157 standards. It is to be noted that 19115-2 includes the bulk of the data quality elements described in 19157, and there are only a few differences which can easily be implemented as add-ons. Also note that the ESDIS Unified Metadata Model (UMM) shows how to apply ISO standards to NASA data. The ESDIS guidance for use of ISO standards [16] states: "Missions will need to work with metadata specialists at their assigned DAACs to meet this requirement. At a minimum, the mission data system should provide enough information that the DAAC can generate sufficient collection level and granule level metadata to ingest into the Common Metadata Repository (CMR). ISO-compliant metadata can be exported from CMR as needed." User resources for ISO metadata, CMR, and the Unified Metadata Model on which CMR is based are provided by reference [16].

3.2. Help Users to Access and Understand Data Quality Information

ESDIS/DAACs should identify and adopt efficient and consistent ways to help data users in accessing and understanding data quality information (e.g., error sources, dataset limitations, and quality assessment) as these would address user questions about data quality and provide user feedback about data quality to user communities. Science teams and PIs should provide guidance to DAACs on including detailed information in product user guides that describes the limitations and/or quality of the data. Examples should be given showing how to apply the quality assessments from the point of view of science teams and/or PIs. ESDIS should identify different ways in which data quality information is currently being conveyed by various DAACs and consolidate these approaches into a consistent mechanism for homogenous, efficient dissemination that results in a more optimal cross-DAAC user experience of data discovery and extraction of data and information about the data. One specific implementation example could be collection level quality information made available in standardized, online guide documents.

3.3. Metadata Authoring and Validating Tools

Metadata is important in conveying information about data, and as such, ESDIS/DAACs should adopt, consolidate, enhance, and/or create metadata authoring and validating tools to assist DAACs, data producers, and data users through the development and validation of richer metadata at collection and granule levels. These tools should also assist data users in validating the metadata. Specifically, these tools should: 1. Support multiple standards, including Unified Metadata Model (UMM) [4], ISO 19115/19157 [6,7] (as implemented by NASA; see section 3.1 above), and Climate and Forecast (CF) [1,14]; 2. Collect minimum required Common Metadata Repository (CMR) [5] and standard-specific metadata; 3. Support population of data quality fields (e.g. the DIF quality field and CF-

defined quality attributes [3]); and 4. Support import/export and translation of CMR metadata with standards-based importable/exportable formats such as XML and JSON.

3.4. Develop Tools to Help Users to Leverage Data Quality Information

ESDIS/DAACs should develop tools to help data users easily use data quality information in their research, such as finding, accessing, and processing data based on user-defined quality criteria. For example, all granule level quality metadata should be accessible through clients such as Earthdata Search [2] and Worldview [15], with the highest-level quality description (e.g., good/bad) prominently displayed alongside granule search results or as a layer in visualization tools. Users should also have access to more detailed granule level quality information (flags, etc.) as an additional filtering mechanism for subsetting and extraction of quality-specified data.

3.5. Finer-levels of Metadata

ESDIS and NASA Earth science programs should support the effort to research and determine required data quality metadata elements at finer-levels (e.g. file, parameter and pixel). Existing quality flag standards for flag attributes should be recommended for datasets where it is appropriate. For example, data formatted using either NetCDF (Version 4 or classic) or HDF (HDF4, HDF5, or HDF-EOSx) should use CF quality flags. Examples of such quality flagging schema are included in DQWG's Data Management Template for Data Processing Systems [9].

3.6. Science Advice

NASA Earth Science Research Program should ensure that each funded project (e.g., MEaSUREs, ACCESS) should adopt a science review board/team to advise data producers on quality and usability of the dataset as it is being developed. Existing review boards (e.g., DAAC User Working Groups) and teams (e.g., NASA science and Cal/Val teams) could be leveraged in this regard but should have oversight to ensure these boards/teams are fulfilling their expectations.

3.7. Facilitate DAAC-PI Communication

ESDIS should develop tools and mechanisms that facilitate communication between DAACs and science teams/PI's to more effectively exchange thoughts and consolidate ideas on describing, representing, and using quality information for data to be created by projects and to be archived at DAACs. Further, the NASA Earth Science Research Program should set policies to facilitate such communication.

3.8. Data Quality Best Practices

DAACs should provide guidance and information on representing data quality as part of data management best practices for data producers to use when developing data and metadata. This should include ensuring that dataset guide documents are created for users. Examples of more general data best practices guidance has been made available by the PO.DAAC [12], ORNL DAAC [17], GHRC DAAC [18], NSIDC DAAC [19], and SEDAC [20].

4. Acknowledgements:

Support by the NASA ESDIS project for the work of the DQWG is appreciated. Additional support was provided by the Earth Science Information Partners (ESIP) through their Information Quality Cluster (IQC) to further evaluate use cases, recommendations and proposed solutions in cooperation with the DQWG as published by Ramapriyan et al. 2017 [13]. We thank reviewers for their constructive comments that improve this document.

5. References

[1] Climate and Forecast Conventions, Latest, 2018: http://cfconventions.org/latest.html (Last accessed February 21, 2019)

[2] Earthdata Search, NASA Earthdata web site, 2018: https://search.earthdata.nasa.gov/ (Last accessed February 21, 2019)

[3] Eaton, B. J. Gregory, B. Drach, K. Taylor, S. Hankin, et al: NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.7, http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.pdf (Last accessed February 21, 2019)

[4] ESDIS Unified Metadata Model, NASA, 2017: https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository/unified-metadata-model-umm (Last accessed February 21, 2019)

[5] ESDIS Common Metadata Repository, NASA, 2017: https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository (Last accessed February 21, 2019)

[6] International Organization for Standards (ISO), 2009: Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data, ISO 19115-2:2009, https://www.iso.org/standard/39229.html (Last accessed February 21, 2019)

[7] International Organization for Standards (ISO), 2013: Geographic information – Data Quality, ISO 19157:2013, https://www.iso.org/standard/32575.html (Last accessed February 21, 2019)

[8] International Organization for Standards (ISO), 2016: Geographic information – Data Quality – Part 2: XML schema implementation, ISO/TS 19157-2:2016, https://www.iso.org/standard/66197.html (Last accessed February 21, 2019)

[9] NASA Earth Science Data Quality Working Group, 2018: Data Management Plan Template for Data Producers, ESDIS Standards Office Technical Note ESDS-RFC-032, https://earthdata.nasa.gov/user-resources/standards-and-references/templates-for-nasa-data-management-plans (Last accessed February 21, 2019)

[10] NASA Earth Science Data Quality Working Group, 2018: Comprehensive Data Quality Recommendations for Data Producers and Distributors, ESDIS Standards Office Technical Note ESDS-RFC-033, https://earthdata.nasa.gov/user-resources/standards-and-references/recommendations-from-the-data-quality-working-group (Last accessed February 21, 2019)

[11] NASA Earth Science Data Quality Working Group, 2018: Solutions Master List for Earth Science Data Quality, Earthdata Wiki, https://wiki.earthdata.nasa.gov/x/2pASBg (Last accessed February 21, 2019)
[12] PO.DAAC Data Management Best Practices, Jet Propulsion Laboratory, California Institute of Technology, Pasadena CA, 2018: https://podaac.jpl.nasa.gov/PO.DAAC_DataManagementPractices (Last accessed February 21, 2019)
[13] Ramapriyan, H., G. Peng, D. Moroni, and C.-L. Shie, 2017: Ensuring and Improving Information Quality for Earth Science Data and Products. D-Lib Magazine, 23. http://dx.doi.org/10.1045/july2017-ramapriyan
[14] Rew, Russ, 2010: CF Metadata Conventions, ESDIS Standards Office Community Standard ESDS-RFC-021, https://earthdata.nasa.gov/user-resources/standards-and-references/climate-and-forecast-cf-metadata-conventions (Last accessed February 21, 2019)
[15] WorldView, NASA Earthdata Web site, 2019: https://worldview.earthdata.nasa.gov (Last accessed February 21, 2019)
[16] ISO 19115 Geographic Metadata Information, NASA Earthdata website, https://earthdata.nasa.gov/user-resources/standards-and-references/iso-19115 (Last accessed February 21, 2019)
[17] ORNL Data Management Best Practices, available online: https://daac.ornl.gov/datamanagement/#best_practices (Last accessed February 21, 2019)
[18] GHRC Data Management Best Practices, available online: https://ghrc.nsstc.nasa.gov/data-publication/ (Last accessed February 21, 2019)
[19] NSIDC Data Management Best Practices, available online: http://nsidc.org/sites/nsidc.org/files/files/data/daac/daac_data_policy_v09-1.pdf (Last accessed February 21, 2019)
[20] SEDAC Data Documentation Template (Example of completed template at https://doi.org/10.7927/H43776SR (Last accessed February 21, 2019)

6. Authors and Contributors

Writing Authors

Zhong Liu (GES DISC)
Hampapuram “Rama” Ramapriyan (SSAI/GSFC ESDIS)
Yaxing Wei (ORNL DAAC)
Chung-Lin Shie (GES DISC)
David Moroni (JPL / PO.DAAC)
Robert R. Downs (SEDAC)
Ted Habermann (HDF Group)
Donna Scott (NSIDC)
George Huffman (GSFC)
Deborah Smith (GHRC)

DQWG Members

[2014-2015]

David Moroni (JPL / PO.DAAC, Chair), Hampapuram Ramapriyan (SSAI, GSFC/ESDIS, Co-Chair), Ed Armstrong (JPL/PO.DAAC), Ross Bagwell (ESDIS), Stacie Doman Bennett (LPDAAC), Charlene DiMiceli (UMD), Feng Ding, Robert R. Downs (SEDAC), Ted Habermann (The HDF Group), Pierre Guillevic (UMD), Steve Olding (ESDIS), Bill Rossow (GSFC), Donna Scott (NSIDC), Suhung Shen (GSFC), Chung-Lin Shie (GES DISC), Marc Simard (JPL), Gilberto Vicente (GSFC), Yaxing Wei (ORNL DAAC)

[2015-2016]

David Moroni (JPL / PO.DAAC, Chair), Hampapuram Ramapriyan (SSAI, GSFC/ESDIS, Co-Chair), Ed Armstrong (JPL/PO.DAAC), Ross Bagwell (ESDIS), Stacie Doman Bennett (LPDAAC), Charlene DiMiceli (UMD), Robert R. Downs (SEDAC), Pierre Guillevic (UMD), Peter Hall (SSAI/GSFC), Molly Hardman (NSIDC), George Huffman (GSFC), Siri Jodha Khalsa (NSIDC), Tiffany Matthews (ASDC), Steve Olding (ESDIS), Donna Scott (NSIDC), Marc Simard (JPL), Chung-Lin Shie (GES DISC), Yaxing Wei (ORNL DAAC)

[2016-2017]

David Moroni (JPL / PO.DAAC, Chair), Hampapuram Ramapriyan (SSAI, GSFC/ESDIS, Co-Chair), Ed Armstrong (JPL/PO.DAAC), Ross Bagwell (SSAI/ESDIS), Stacie Doman Bennett (LPDAAC), Michelle Butler (NCSA), Charlene DiMiceli (UMD), Larry Di Girolamo (UIUC), Robert R. Downs (SEDAC), Yonsook Enloe (SSAI/ESO), Pierre Guillevic (UMD), Ted Habermann (The HDF Group), Peter Hall (SSAI/GSFC), Molly Hardman (NSIDC), Beth Huffer (LaRC), George Huffman (GSFC), Siri Jodha Khalsa (NSIDC), Shannon Leslie (NSIDC), Andrew Mitchell (ESDIS), Steve Olding (SSAI/ESDIS), Byron Peters (SSAI/ESDIS), Donna Scott (NSIDC), Deborah Smith (GHRC DAAC), Chung-Lin Shie (GES DISC), Yaxing Wei (ORNL DAAC), Greg Yetman (SEDAC)

[2017-2018]

Yaxing Wei (ORNL DAAC, Chair), David Moroni (JPL/PO.DAAC, Chair), Hampapuram Ramapriyan (SSAI, GSFC/ESDIS, Co-Chair), Ed Armstrong (JPL/PO.DAAC), Ross Bagwell (ESDIS), Michelle Butler (UIUC), Charlene DiMiceli (U. of Maryland), Feng Ding (GSFC/GES DISC), Robert R. Downs (SEDAC), Carolyn Gacke (LP.DAAC), Larry Di Girolamo (UIUC), Scott Gluck (JPL), Pierre Guillevic (U. of Maryland), Lindsey Harriman (LP.DAAC), Ted Habermann (The HDF Group), Beth Huffer (ASDC), George Huffman (NASA Goddard Space Flight Center), Wenhao Li (JPL), Zhong Liu (GSFC/GES DISC), Sydney Neeley (LP.DAAC), Steve Olding (SSAI/ESDIS), Byron Peters (SSAI/ESDIS), Donna Scott (NSIDC), Suhung Shen (GSFC/GES DISC), Chung-Lin Shie (GSFC/GES DISC), Deborah Smith (GHRC DAAC), James Tilton (GSFC)

Use Cases Contributors

Ed Armstrong (JPL/PO.DAAC), Stacie Doman Bennett (LPDAAC), Lisa Booker (NSIDC), Chris Derksen (U. of Waterloo), Feng Ding (GSFC), Jessica Hausman (JPL/PO.DAAC), Nathan Kurtz (GSFC), Christopher Lynnes (ESDIS), David Moroni (JPL/PO.DAAC), Hampapuram Ramapriyan (SSAI, GSFC/ESDIS), Marc Simard (JPL), Vardis Tsontos (JPL/PO.DAAC)

APPENDIX A – Glossary of Acronyms and Abbreviations

ACCESS	Advancing Collaborative Connections for Earth System Science
CF	Climate and Forecast (metadata conventions)
CMR	Common Metadata Repository
CSDGM	Content Standard for Digital Geospatial Metadata
DAAC	Distributed Active Archive Center
DQWG	Data Quality Working Group
EOSDIS	Earth Observing System Data and Information System
ESDIS	Earth Science Data and Information System (Project)
ESDS	Earth Science Data System (Program)
ESDSWG	Earth Science Data Systems Working Groups
FGDC	Federal Geographic Data Committee
GES DISC	Goddard Earth Sciences Data Information Services Center
GHRC	Global Hydrology Resource Center
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
HQ	Headquarters
ISO	International Organization for Standardization
JPL	Jet Propulsion Laboratory
LaRC	Langley Research Center
LHF	Low-Hanging Fruit
LPDAAC	Land Processes Distributed Active Archive Center
MEaSUREs	Making Earth System Data Records for Use in Research Environments

NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputing Applications
NetCDF	Network Common Data Form
NSIDC	National Snow and Ice Data Center
ORNL DAAC	Oak Ridge National Laboratory Distributed Active Archive Center
PI	Principal Investigator
PO.DAAC	Physical Oceanography Distributed Active Archive Center
SDS	Science Data System
SEDAC	NASA Socioeconomic Data and Applications Center
SIPS	Science Investigator-led Processing System
SSAI	Science Systems and Applications, Inc.
UIUC	University of Illinois Urbana-Champaign
UMD	The University of Maryland
UMM	Unified Metadata Model

Appendix B – Solutions Master List Summary

No.	Solution Name	Solution Summary (used to derive relevance)	Implementation Strategy	Benefits of Proposed Implementation Solutions	Solution Point of Contact	Reference URLs
1	Collaboratory for quAlity Metadata Preservation (CAMP) - ASDC	CAMP is currently being developed and expanded upon for the ASDC metadata reconciliation efforts. As development progresses, the ASDC will leverage this platform as a centralized repository for metadata entry/revisions, new data submission requests, and interoperability for both internal (i.e. OPenDAP) and external (i.e. CMR REST API) systems to streamline metadata management and increase transparency for the data ingest process. The end goal is to provide a UI for direct metadata entry by ASDC members and data providers. Validate CMR Compliance. Metadata for current ASDC (BEDI identified) data products has been imported into the CAMP Database and are within weeks of being validated by science teams. Depending on required CMR fields, there may be fields added.	Facilitate DAAC - PI Communication; Support Metadata Creation (dataset-level)	1. Confidence in metadata accuracy 2. Quick and easy to provide metadata 3. Metadata completeness	Tiffany Trapasso (tiffany.j.mathews@nasa.gov)	https://camp.larc.nasa.gov/
2	Metadata Compliance Checker (PO.DAAC)	Provides tool for both DAACs and Data Producers to evaluate metadata standards compliance at granule level. Multiple forms of compliance check: ACDD, CF, … quality flags, completeness/compliance, … netCDF/HDF/OPeNDAP, Target at data producers as major user community. Output report from the checker will contain useful information and be exposed to end users? validate time against ISO 8601	Standards Compliance Checking and Reporting (granule-level); Support Metadata Creation (granule-level)		David Moroni (David.F.Moroni@jpl.nasa.gov)	http://podaac-uat.jpl.nasa.gov/mcc/
3	ATRAC (NOAA/NCEI/NCDC)	Provides open web form for metadata entry by data producer which is interfaced with a backend metadata archive database maintained by the data center. Note: Whether this resource/tool is developed directly by ESDIS or by a DAAC, the important aspect is that the DAAC must have immediate access to the metadata that is input by this tool for the purpose of verifying accuracy and completeness.	Support Metadata Creation (dataset-level); Standards Compliance Checking and Reporting (dataset-level);		David Moroni (David.F.Moroni@jpl.nasa.gov)	https://www.ncdc.noaa.gov/atrac/index.html
4	ORNL DAAC Ingest Automation System	Tool is developed by ORNL DAAC and provides a more automated workflow for data submissions intended to increase efficiency of DAAC/Producer communications regarding new datasets or new versions of datasets. Tool could be optimized or extended to include additional information exchange for data quality and or quality flag information. Core functions include: 1) Track data ingest; 2) Automate ingest; 3) Streamline communication; 4) Central management system	Facilitate DAAC - PI Communication; Support Metadata Creation (dataset-level);		David Moroni (David.F.Moroni@jpl.nasa.gov)	https://git.earthdata.nasa.gov/projects/DAACSUB/repos/daac-ingest-dashboard/browse ORNL DAAC Ingest Automation Swimlanes: https://wiki.earthdata.nasa.gov/download/attachments/54330067/ornl_daac_ingest_automation_swimlanes.png?version=1&modificationDate=1447171490883&api=v2 Presentation at 2015 ESIP Summer Meeting: http://live.commons.esipfed.bluedotapps.org/sites/default/files/6_Wright-Vannan_ORNL_DAAC_ingest_071515.pptx
5	Ocean CO2 Metadata Collection Form	Collection-level metadata collection form developed by ORNL for oceanic in situ observation datasets tailored for CO2 collection. Could potentially be extended to include satellite datasets. Same as the metadata editor in the ORNL DAAC Ingest Automation System	Metadata creation support (dataset-level)		David Moroni (David.F.Moroni@jpl.nasa.gov)	http://mercury.ornl.gov/OceanOME/
6	Data Quality Guide Document	A standardized template document design to provide users with familiar and comparable data quality guidance for all data sets sharing a common measurement parameter. Data quality templates for MEaSUREs to fill out.	Guidance, Instruction, and Dissemination (for data users)		David Moroni (David.F.Moroni@jpl.nasa.gov) Hampapuram Ramapriyan (hampapuram.ramapriyan@ssaihq.com)	A few examples - found by Google search for AIRS, CERES and MODIS Data Quality. 1. http://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/AIRS/3.3_ScienceDataProductDocumentation/3.3.5_ProductQuality/V6_L2_Quality_Control_and_Error_Estimation.pdf 2. https://eosweb.larc.nasa.gov/project/ceres/quality_summaries/CER_SSF_Terra_Edition3A.pdf 3. http://ceres.larc.nasa.gov/dqs.php 4. https://lpdaac.usgs.gov/sites/default/files/public/modis/docs/MODIS_LP_QA_Tutorial-1b.pdf 5. https://globalmonitoring.sdstate.edu/sites/default/files/QA_paper.pdf 6. http://modis-atmos.gsfc.nasa.gov/_docs/QA_Plan_C6_Master_2015_05_05.pdf
7	ACT-America Science Data Working Group	A Science Data WG, including participants (funded by the project) from data centers (ORNL DAAC and ASDC) and different research groups, was formed in the ACT-America project to 1) coordinate data management activities with instrument teams, modelers, remote sensing, and external data sources and 2) ensure data, products, and information required to address science questions are available in harmonized forms when needed. Telecons are held periodically to exchange any data-related thoughts between research groups and the data centers. Currently solution is applicable to modeling (ORNL DAAC) and Airborne observations (ASDC) components of ACT-America data management; But can be applicable to others.	Facilitate DAAC - PI Communication	1. Coordinate data management activities with instrument teams, modelers, remote sensing, and external data sources 2. Ensure data, products, and information required to address science questions are available in harmonized forms when needed.	Yaxing Wei (weiy@ornl.gov) Stacie Doman Bennett (sdomanb@gmail.com)	N/A

8	(NASA) Science Advisory Team	A NASA assigned team to review data for each project/product, such as "NASA SAT MEaSUREs WELD". These scientists would be assigned to a project/product team, are recognized as experts in the specified field(s), and serve to advise the verification and quality of final distributed products.	Data Quality Information (science perspective); Guidance, Instruction, and Dissemination	1. Provides early adopters to data products from NASA Earth Science remote sensing projects. 2. Provides beta testers for MEaSUREs ESDRs.	Stacie Doman Bennett (sdomanb@gmail.com)	
9	Data Quality Section in Data Management Plan	Recommend including a section on data quality in the Data Management Plan to be created for each project, such as MEaSUREs, after the award, as a living document to be updated as more details about the data are identified. (It is possible that the initial version of the DMP is prepared before details are known, since it is to be delivered early in the project).	Guidance, Instruction, and Dissemination (for data producers and DAACs); Facilitate DAAC - PI Communication; Dissemination on Data Quality Information;		Hampapuram Ramapriyan (hampapuram.ramapriyan@ssaihq.com)	http://science.nasa.gov/media/medialibrary/2012/05/07/Data_Mgmt_Plan_guidelines-20110111.pdf document has brief comment in section 3.2.3: "This section should also describe project requirements and plans for assuring and documenting data quality including validation and release of products to the archive system." HS3 and CARVE DMP's have material on data quality while AirMOSS DMP does not.
10	DAACs DMP (or Data Management Guidlines)	Some DAACs (e.g., PODAAC,SEDAC, ...) write their own DMPs for specific datasets or a collection of datasets for the purpose of managing datasets throughout their lifecycle. PO.DAAC is currently finalizing a standardized template for the DAAC-specific DMP. The SEDAC Data Nomination template is used internally and contains sections to capture data quality information. The ORNL DAAC doesn't have DMPs for specific datasets. Instead, it provides general guidance for data providers to conduct data management and prepare for data archival.	Guidance, Instruction, and Dissemination (for data producers and DAACs); Facilitate DAAC - PI Communication		David Moroni (David.F.Moroni@jpl.nasa.gov) Yaxing Wei (weiy@ornl.gov) (knowledge authority)	ORNL DAAC Data Management Guidance for Data Providers: http://daac.ornl.gov/PI/pi_info.shtml
11	Kayako	Several DAACs have integrated Kayako, a customer service software, into their Websites to replace old ways of conducting user support. User questions and feedbacks for different DAACs are now managed consistently.	User Services (Help Desk); Knowledgebase (for data users)	Kayako provides an integrated system for ESDIS and individual DAACS to easily track and coordinate user questions and feedbacks related to data products, websites, tools, etc. It also allows individual DAACs to easily compile knowledge bases and FAQs by pulling past user support records from Kayako system.	Yaxing Wei (weiy@ornl.gov)	Contact US on http://daac.ornl.gov/ and http://daymet.ornl.gov/ https://support.earthdata.nasa.gov
12	Daymet Website	The ORNL DAAC developed a project website dedicated for Daymet: http://daymet.ornl.gov . It is different from the landing pages of Daymet data sets. This website provides information about Daymet data description, documentation, visualizations, data access tools and services, publications using Daymet data, Daymet-related tools contributed by the users community, and news update.	Data quality information (program-specific collection);	Daymet website can be considered as one way to convey data product, including quality, information to data users. Daymet is becoming probably the most popularly used data product recently. The Daymet website helps a lot, even though it's hard to quantify its impact on this popularity.	Yaxing Wei (weiy@ornl.gov)	http://daymet.ornl.gov/
13	Identify different ways in which DAACs are conveying data quality information	Identify different ways in which data quality information (e.g. quality flags and known issues) is being conveyed by various DAACs. Understand why they need to be different. To the extent possible arrive at common approaches. At least a minimal common set of items should be shown on data quality pages at the DAACs.	Data quality information (dataset-level); Different approaches for data quality information (dataset-level)	Although most user guides contain some information on data quality, it would be good to provide guidance so that it is consistent and complete as possible.	Hampapuram Ramapriyan (hampapuram.ramapriyan@ssaihq.com)	
14	FAQ Development and Analysis (UserVoice)	Populate a set of FAQs for each new data set upon release by anticipating possible questions that users might ask. From FAQ, identify data sets receiving excessive questions as those to be considered for dissemination of additional or enhanced documentation. SEDAC example	User Services (Help Desk); Knowledgebase (for data users)			http://sedac.uservoice.com/knowledgebase
15	NASA GSFC Data Quality Screening Service	A tool developed by Christopher Lynnes & Richard Strub for GES-DISC. "DQSS is designed to screen data using both ontology based criteria and user selections of quality criteria (such as minimal acceptable QualityLevel). Data that do not pass the criteria are replaced with fill values, resulting in a file that has the same structure and is usable in the same ways as the original." This service can be utilized before data ingest for the distributor. This service can also be utilized by the public - to further screen the product's quality.	Data quality screening (granule-level filtering)	Provides DAACs a tool to understand quality attributes for overall documentation to product validation. Provides Users a tool to better understand how data decisions regarding quality were established.	Stacie Doman Bennett (sdomanb@gmail.com)	http://opensource.gsfc.nasa.gov/projects/DQSS/
16	CF granule metadata	Implementation of CF Conventions for quality variables to require flag_values, flag_mask, flag_meanings CF attributes	Guidance and instruction; Data quality and information		Ed Armstrong (edward.m.armstrong@jpl.nasa.gov)	http://cfconventions.org/
17	Document Error Sources/Limitations/ Quality Assessment	Provide guidance to DAACs on including detailed information in product user guides that describes the limitations &/or quality of the data	Data quality and information; User Services	Although most user guides contain some information on data quality, it would be good to provide guidance so that it is consistent and complete as possible.	Donna Scott (dscott@nsidc.org)	http://nsidc.org/data/docs/daac/smap/sp_l2_smp/index.html#errorsource (would be good to get examples from other DAACs)

18	LP DAAC Project Lifecycle Plan (PLP)	This document is written from the point of view of the LP DAAC, advocators for products as they move through the lifecycle from Inception to Active Archive to Long Term Archive, and advocators of products that adhere to interoperability standards. Product capture is the first step in providing community-wide access to data and information. PO.DAAC has a very similar policy that covers a series of project lifecycle planning documents and artifacts known as the "Dataset Lifecycle Policy".	Guidance and instruction	DAAC Scientist is part of the NASA funded dataset development - with focus on guidance and communication from the project start.	Stacie Doman Bennett (sdomanb@gmail.com)	http://pubs.usgs.gov/of/2014/1139/pdf/ofr2014-1139.pdf http://podaac.jpl.nasa.gov/PO.DAAC_DataManagementPractices#Dataset%20Lifecycle
19	EUFAR Metadata Creator	Online metadata authoring tool that creates INSPIRE-compliant metadata in XML for the EU Facility for Airborne Research. But only free text for quality input.	Metadata creation support	Facilitates entry of metadata and produces output that is standards compliant in content and format.	SiriJodha Khalsa (sjsk@nsidc.org)	http://176.31.165.18:8080/emc-eufar/
20	ISO Data Quality elements	A webpage describing elements of the ISO 19157 data quality metadata standard	Guidance and instruction; Metadata creation support	Until there is a NASA profile of the ISO metadata standard, metadata authors need guidance on how to express quality in ISO. This provides a guide.	SiriJodha Khalsa (sjsk@nsidc.org)	https://geo-ide.noaa.gov/wiki/index.php?title=ISO_Data_Quality ECHO Data Quality Metadata in ISO: https://wiki.earthdata.nasa.gov/display/ESDSWG/ECHO+Data+Quality+Metadata+in+ISO
21	schema for ISO metadata, including Data Quality	zip file containing schema for all 19115 and related metadata ISO standards	Metadata creation support	If authoring metadata conforming to ISO standards (without a tool, or in customizing an existing tool) one need the schema for the standard.	SiriJodha Khalsa (sjsk@nsidc.org)	http://standards.iso.org/iso/19115/19115.zip
22	NCO Utilities for granule level metadata authorship, editing, and standardization	allows addition/modification of quality attributes in netCDF files	Metadata creation support	Facilitates creation and modification of metadata that complies with CF conventions. Specific to netCDF and HDF. Being expanded under EarthCube award "Advancing netCDF-CF for the Geoscience Community"	David Moroni (David.F. Moroni@jpl.nasa.gov)	http://nco.sourceforge.net/
23	AADC Metadata XML conversion script	py script that loops over metadata DIF XML files and converts them to other XML formats using XSL files.	Metadata creation support	This script would be useful for converting existing GCMD DIF records to, e.g. ISO.	SiriJodha Khalsa (sjsk@nsidc.org)	https://github.com/AustralianAntarcticDataCentre/metadata_xml_convert
24	PO.DAAC User Forums	The PO.DAAC has established a user forum to service user inquiries on all data issues including data quality concerns. This forum is URS-compliant and also provides the ability to directly create a Kayako ticket for timely help desk support.	User Services (Help Desk); Knowledgebase (for data users)	Provides FAQ's, data recipes, discussions on data quality issues, and discipline-specific discussion threads.	David Moroni (David.F. Moroni@jpl.nasa.gov)	https://podaac.jpl.nasa.gov/forum/
25	Virtual Quality Screening Service	Provides an interface to screen L2/L3/L4 SMAP and GHRSST physical retrieval observations using quality information (variables) contained within the granules. Provides a data extraction method once the quality screening filters have been defined. Returns only the quality filtered data.	Data Quality Information Representation; Guidance, Instruction, & Dissemination	User extracts only the data that meets their quality specifications set using quality flags, bit flags, or other variables.	Ed Armstrong (edward.m. armstrong@jpl.nasa.gov)	http://podaac-access.jpl.nasa.gov
26	MODIS Python Toolbox for ArcGIS	Data values in MODIS quality layers are store as bit-packed integer values. To get at the information stored in the data values, users must first converted the integer value to its binary representation then interpret each specified bit combinations (bit words) which characterize particular quality attributes. The MODIS Python Toolbox contains a tool (DecodeQuality) that decodes MODIS quality layers, and returns individual thematic GeoTIFFs for each quality attribute.	Data quality information	Provides thematic GeoTIFFs for each quality attribute contained in the original bit-packed data value.	Aaron Friesz (aaron. friesz.ctr@usgs.gov)	https://git.earthdata.nasa.gov/projects/LPDUR/repos/arcgis-modis-python-toolbox/browse