

Data Management Plan Template for DAACs

Status of this Memo

This document introduces a revised template for Data Management Plans (DMPs), which may be developed as needed by the EOSDIS Distributed Active Archive Centers (DAACs). The version of the DMP template for DAACs dated September 2016 is updated to include recommendations from the Data Quality Working Group (DQWG). The distribution of this document is unlimited.

Change Explanation

Not applicable. This RFC is not an update to an existing RFC, but a revision to a document that was produced by the ESDIS Project in September 2016 and distributed to the DAACs.

Copyright Notice

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Abstract

This document is derived from recommendations developed during 2014-2016 by the Data Quality Working Group (DQWG), one of NASA's Earth Science Data System Working Groups. The recommendations result from an analysis of use case studies that highlight issues regarding capturing, describing and conveying information about the quality of data products held at EOSDIS DAACs. The Data Management Plan Template for DAACs incorporating the DQWG recommendations is attached to this RFC as Appendix B, and is available as an independent document.

1 Introduction

This document is derived from the recommendations developed during 2014-2016 by the Data Quality Working Group (DQWG), which is one of NASA's Earth Science Data System Working Groups (ESDSWG). These recommendations were a result of the analysis of 16 use cases describing users' explorations of selected data products held at the EOSDIS DAACs. These use cases highlighted issues and needs regarding capturing, describing and conveying the information about the quality of data products to users such that they can access the information, understand the data quality, and properly use the data products. The members of the DQWG made recommendations for improvement, some of which apply to data producers and others to data distributors (DAACs). These recommendations [1,2] are published through ESDIS Standards Office (ESO), which should provide greater visibility into the origin and rationale behind most of what is being proposed in this document. [The term "data producers" is used in this document to refer to science teams that develop algorithms as well as people managing/running the data production systems (e.g., Science Investigator-led Processing Systems - SIPSs) who work closely with the algorithm developers.] The recommendations applicable to the DAACs were used for updating the "Data Management Plan template for

DAACs" that had been developed by the ESDIS Project in September 2016 and distributed to the DAACs at that time. DAACs should use this updated template when they make updates to their DMPs. A separate document [3] is developed for the data producers to ensure that they include material addressing data quality in their DMPs. It should be noted that much of what the DAACs might address in their respective DMPs will be carried over from existing DMPs previously drafted by data producers. While the DQWG's focus is on the Data Quality section, other sections have also been updated to account for the DQWG recommendations since capturing and conveying information on data quality also involves data stewardship, data discovery and access.

2 RFC Content

Please see Appendix B: "Data Management Plan Template for DAACs".

3 References

[1] NASA Earth Science Data Quality Working Group, Comprehensive Data Quality Recommendations for Data Producers and Distributors, ESDIS Standards Office Suggested Practice Technical Note ESDS-RFC-033, <https://earthdata.nasa.gov/user-resources/standards-and-references/recommendations-from-the-data-quality-working-group>.

[2] NASA Earth Science Data Quality Working Group, High-Priority Data Quality Recommendations for Data Producers and Distributors, ESDIS Standards Office Suggested Practice Technical Note ESDS-RFC-034, <https://earthdata.nasa.gov/user-resources/standards-and-references/recommendations-from-the-data-quality-working-group>.

[3] NASA Earth Science Data Quality Working Group, Data Management Plan Template for Data Producers, ESDIS Standards Office Convention Specification ESDS-RFC-032, <https://earthdata.nasa.gov/user-resources/standards-and-references/templates-for-nasa-data-management-plans>.

4 Authors' Addresses

This document was prepared by the DMP DQ Section Subgroup of the NASA Earth Science Data Systems Working Groups' Data Quality Working Group. The names of the members of this subgroup are listed below:

H. K. "Rama" Ramapriyan - Science Systems and Applications, Inc. and Earth Science Data and Information System Project, NASA Goddard Space Flight Center, Tel: 805-402-7125, Hampapuram.Ramapriyan@ssaihq.com (Chair, DMP DQ Section Subgroup)

Donna Scott - National Snow and Ice Data Center (Co-Chair, DMP DQ Section Subgroup)

Ed Armstrong - Physical Oceanography DAAC, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Charlene DiMiceli - University of Maryland

Robert Downs -Socio-Economic Data and Applications Center, Columbia University

Carolyn Gacke - Land Processes DAAC

Scott Gluck - Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

George Huffman - NASA Goddard Space Flight Center

Zhong Liu - NASA Goddard Earth Science Data and Information Service Center (GES DISC)

David Moroni - Physical Oceanography DAAC, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Chung-Lin Shie - NASA Goddard Earth Science Data and Information Service Center (GES DISC)

Yaxing Wei – Oak Ridge National Laboratory DAAC

Appendix A - Glossary of acronyms and abbreviations

Acronym / Abbreviation	Description
DAAC	Distributed Active Archive Center
DMP	Data Management Plan
DQ	Data Quality
DQWG	Data Quality Working Group
EOSDIS	Earth Observing System Data and Information System
ESD	Earth Science Division
ESDIS	Earth Science Data and Information Systems
ESDSWG	Earth Science Data System Working Groups
ESO	ESDIS Standards Office
NASA	National Aeronautics and Space Administration
RFC	Request For Comment
SIPS	Science Investigator-led Processing System

Appendix B -- Data Management Plan Template for DAACs

Each organization funded by NASA to produce science data is required to prepare a DMP at the time it is funded and maintain it as a living document by reviewing it periodically and ensuring that it is up-to-date. The requirements calling for DMPs are documented in appropriate sections of NASA Procedural Requirements (NPR) 7120.5 [1] and 7120.8 [2]. The former applies to Space Flight Programs and Projects, and the latter to Research and Technology Programs and Projects. In addition, NASA responded in February 2014 [3] to the Executive Office of the President's Office of Science and Technology Policy (OSTP) Memorandum ([4], dated February 22, 2013) for the Heads of Executive Departments and Agencies, "Increasing Access to the Results of Federally Funded Scientific Research", with a plan titled "NASA Plans for Increasing Access to the Results of Scientific Research – Digital Scientific Data and Peer Reviewed Publications". This plan covers the requirements for NASA funded researchers – "...all extramural researchers receiving NASA grants, cooperative agreements, and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for the long-term preservation and access of scientific data in digital format." To apply these principles with EOSDIS components, the ESDIS Project has developed templates tailoring the DMPs to the specific types of activities the components perform.

Data producers should provide a DMP covering their products to the DAACs. DAACs may also choose to develop more general DMPs to describe their data operations. The purpose of this template is to provide DAACs guidance on the contents of Data Management Plans (DMPs). The DAAC DMPs are prepared by the respective DAACs, reviewed by the ESDIS Deputy Project Manager for Operations and approved by the ESDIS Project Manager. The DMPs are valid for 2 years and updated as needed (either as a result of a 2-year review or "off-cycle" changes that may be needed). In addition, DAACs may update DMPs periodically with new and more robust information as the missions and projects commissioned with producing the data are maturing. DAAC DMPs will be maintained as configuration controlled documents by the ESDIS Project.

Figure 1 depicts the DMP process flow. Note that the review and update of a DMP can be initiated at any time it is deemed necessary. The decision indicated by the rhombus 1.2 implies a requirement to review and update the DMP every 2 years (the "Yes" decision in 1.2), but provides a path for an ad hoc review (the "No" decision in 1.2) and modification of the DMP if necessary.

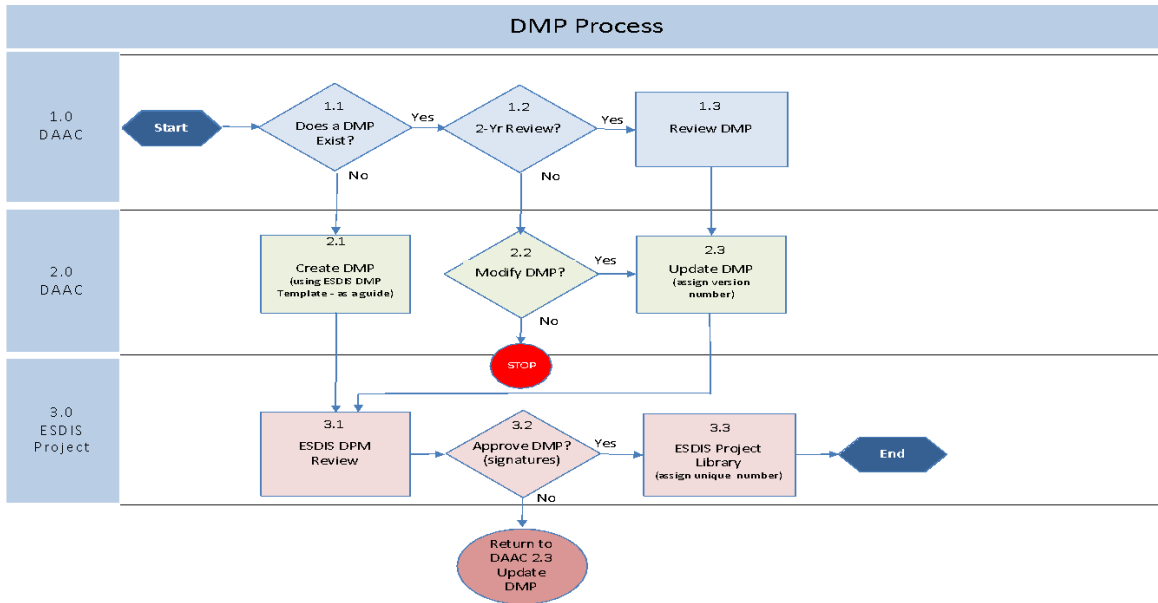


Figure 1: DAAC DMP Process Flow

1. Introduction

Brief description of what the DAAC does and what types of data (Earth science) it archives and distributes.

1.1 Purpose and Scope (of DMP)

For example: “The purpose of this DMP is to describe the [NAME] DAAC's plans for archiving, distribution and provision of user services for Earth science datasets assigned to the DAAC in [discipline(s)].”

1.2 DMP Development, Maintenance and Management Responsibility

For example: “The [NAME] DAAC is responsible for the development, maintenance, and management of the DMP. The [NAME] DAAC Manager, [DAAC manager's name], has overall responsibility for the plan, and has specific responsibility for approving any changes to the plan. All changes to the DMP will be controlled.”

2. Data Accession and Deaccession Overview

Include an opening paragraph in this section describing the process by which datasets are assigned/accepted for archival in the DAAC. The paragraph may use the parts of the wording below as applicable to the DAAC.

“Most datasets are assigned to the DAAC by NASA HQ through instructions to the ESDIS Project. These datasets may be generated by individual instrument science teams from satellite missions or aircraft investigations, field experiments, or PI-led projects (i.e. MEaSUREs). Such assignments of datasets are documented through the ESDIS Project’s Planning, Programming, Budgeting, and Execution (PPBE) process. Some datasets are submitted by data producers with request for archiving at the DAAC. Such datasets are accepted for archiving using the process illustrated in Figure 2. When a dataset is to be deaccessioned or purged, the DAAC will follow procedures outlined in CEOS Data Purge Alert Procedure [5], which provide for orderly removal of datasets from archives giving opportunities for other archives to take on the responsibility for continued archival if deemed appropriate. (Deletion of older versions of datasets that are superseded by newer versions at the direction of the PI/science teams is not considered deaccession.)”

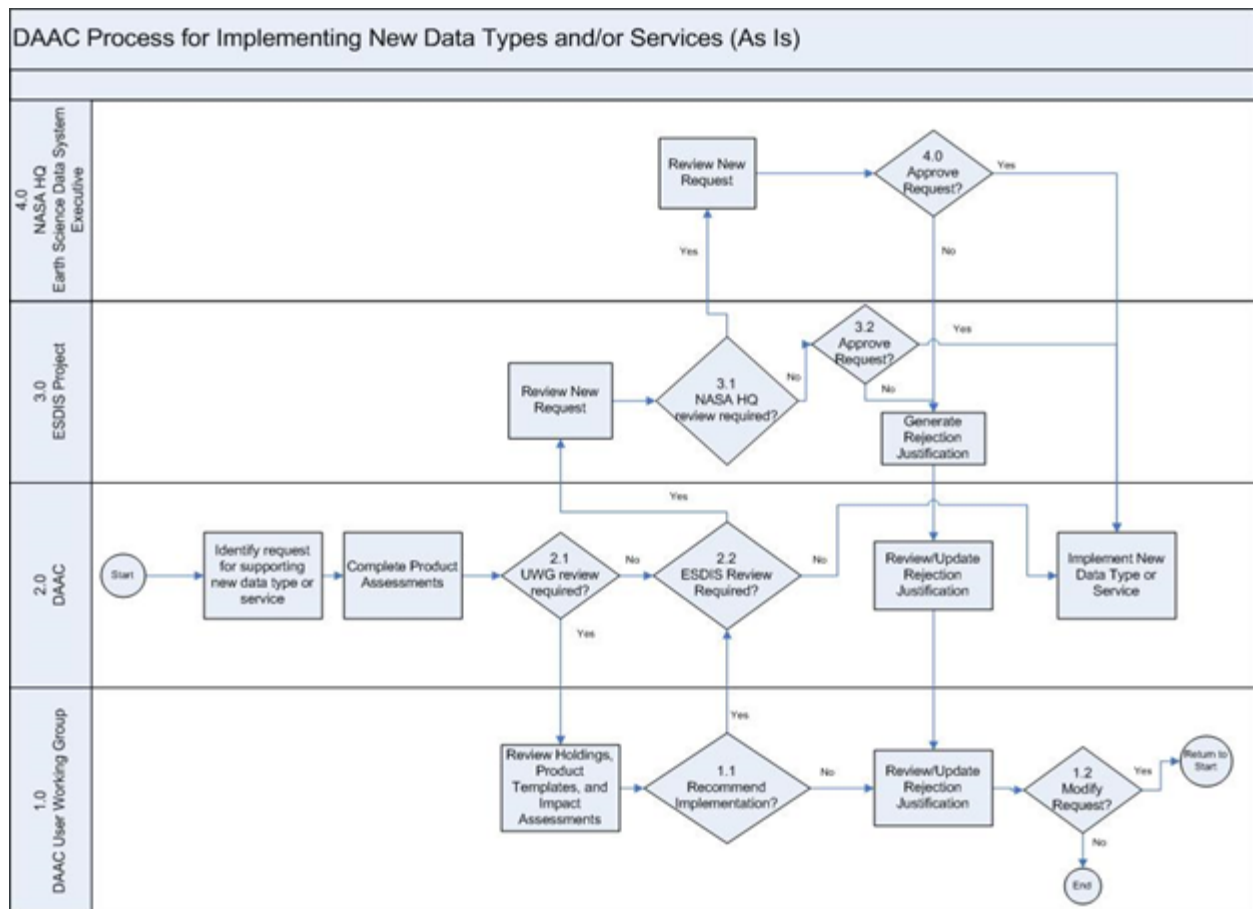


Figure 2. DAAC Process for Implementing New Data Types and/or Services

3. Data Sources Overview

Include a list of all the sources (e.g., missions, instruments, projects or investigations) for whose data the DAAC is responsible. For each case provide a brief descriptive section, with pointers to more detailed information (e.g., mission DMPs, project data repository descriptions). Ensure that the information contained here matches what is in the [EOSDIS Missions Google sheet](#)¹ (update the sheet as needed).

The initial version of the DMP should include all the sources of data the DAAC is currently responsible for. In subsequent versions, the additional sources can be documented in appendices to the DMP.

4. Interfaces

Provide a diagram showing all the major interfaces that are needed for the DAAC to perform its data management functions. Examples of entities with which the DAAC may have interfaces are: data producers (SIPs, Mission Science Data Systems), EOS Data and Operations System (EDOS), ESDIS Project, and other DAACs. Provide pointers/references to the interface control documents or other agreements that specify and describe the interfaces. A brief description of the purposes of the interfaces should be included here.

5. Data Processing

If the DAAC performs any data processing to generate standard products, briefly describe the processing activities here. Include: description of processing system, data flow diagram(s) to support processing, source of ancillary data needed for processing, source of processing software, software integration and testing process, configuration control, and operational procedures. Also see Data Management Plan Template for Data Producers [7]. If there are other documents describing some or all of these items, point to such documents and summarize here.

6. Data Quality

Describe aspects of data quality as it relates to the creation and use of datasets provided to the DAAC for archival and distribution. The aspects of data quality map to four focus areas– 1) Accuracy, Precision and Uncertainty, 2) Distinguishability, 3) Applicability, and 4) Usability [6]. In general, the provision of information on data quality is the responsibility of the PIs and/or science teams responsible for generating the data. These data providers are expected to develop DMPs following the Data Management Plan Template for Data Producers [7], which includes recommendations for contents of the section on data quality.

¹ <https://docs.google.com/spreadsheets/d/1rGRw1cFZkss-UUopoFICjAfiniXNopvxFgfFMuSoZ78/edit#gid=0>

6.1 Capturing (*deriving, collecting and organizing the information*)

6.1a Identify the format(s) chosen for the datasets archived at the DAAC. Data formats should follow NASA ESDIS standard formats as described in [Standards, Requirements and References](#)² where applicable. (NOTE: The intent here is not to get a list of all the datasets and their formats, but just the formats chosen for the DAAC's holdings as a whole.) Describe how the DAAC encourages/ensures that the data producers providing the data will follow the standard formats. List any "best practices" documents developed for the data producers that show how standard formats are used for their datasets, periodic meeting for providing advice to data providers, etc.

6.1b DAACs may assist data producers by providing information on best practices relevant to specific instruments/measurements and verifying data quality attributes of a dataset or data collection. Describe the DAAC process to provide such assistance to data producers. Data producers are responsible for providing data quality flags and/or indicators as well as descriptive data quality documentation. Data producers are responsible for documenting their processes to accomplish these as part of their Data Management Plans for Data Producers.

7. Data Stewardship

7.1 Ingest and Levels of Service Planning

DAACs should identify how they will interface with the data producers to stay informed about their plans and schedules for production phases and reprocessing, including interim, operational, and post-operational phases. These delivery phases can impact the [data maturity levels](#).³ Furthermore, there should be clear distinctions on the levels of service that should be provided in conjunction with these data designations, and to ensure that those planned levels of service are consistent with NASA's [open data policy](#).⁴

7.2 Prevention of loss of data

Describe the plans and processes including reconciliation procedures used by the DAAC to ensure that the data are archived with no loss. Include or point to data migration plans and back-up and recovery strategies. [NOTE: As DMPs are produced, it would be a good time for the DAACs to review the respective Risk Assessment Code (RAC) matrices that show Data Loss Risk and User Impact, and make sure they are up-to-date. These RAC matrices will be maintained by the ESDIS Project separately from the DMPs.]

7.3 Metadata

Identify the metadata standard(s) suitable for the datasets archived at the DAAC. Outline the metadata needs to support data discovery, data preservation, and data use. Metadata formats, at

² <https://earthdata.nasa.gov/user-resources/standards-and-references>

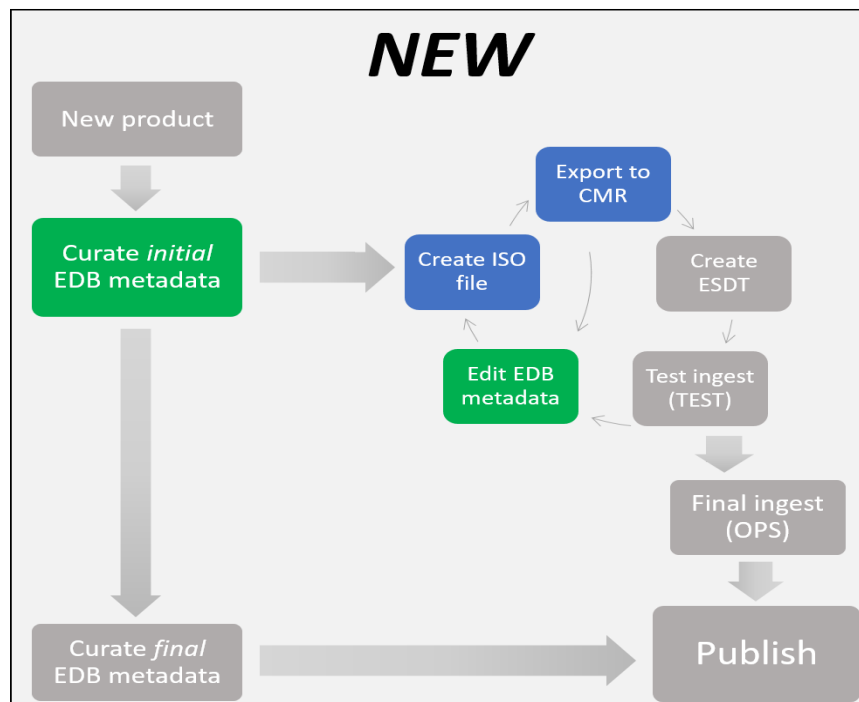
³ <https://science.nasa.gov/earth-science/earth-science-data/data-maturity-levels/>

⁴ <https://earthdata.nasa.gov/earth-science-data-systems-program/policies/data-information-policy>

all different levels (e.g. collection and granule), should follow NASA ESDIS standards as described in [Standards, Requirements and References](#)⁵.

7.3a Provide information about any metadata authoring and validation tools that are planned to be used to ensure that the delivered datasets have useful and complete metadata. For example, the [PO.DAAC Metadata Compliance Checker](#)⁶ as used for granule-level metadata validation (CF, ACDD, and ISO-8601) or NASA's [Common Metadata Repository](#),⁷ for both validation and cataloguing of entire metadata collections.

7.3b Include plans and processes for providing metadata to CMR. For example, a workflow diagram similar to this NSIDC DAAC workflow is a sufficient way to show the metadata to CMR process.



7.4 DOIs and Dataset Information Pages (or Landing Pages)

Include plans and processes for establishing DOI's for datasets, maintaining dataset information pages (or landing pages), and ensuring that datasets offered by the DAACs can be conveniently cited by users in their publications.

⁵ <https://earthdata.nasa.gov/user-resources/standards-and-references>

⁶ <https://podaac-uat.jpl.nasa.gov/mcc/>

⁷ <https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository>

7.4a DOIs – DAACs should assign DOIs in the manner that is consistent with the current [NASA DOI process](#)⁸ described on the EarthData Wiki.

7.4b Landing Pages – Describe how the DAAC implements the corresponding dataset landing page or informational pages. For example, “The [NAME] DAAC utilizes internal databases to dynamically generate a dataset landing page that includes the dataset authoritative ID, data citation, DOI and data access.”

7.5 Collection of Associated Preservation Content

Include plans and processes used for collecting software, documentation, etc., as called for in the Preservation Content Specification [8] to ensure that the contents are captured before the end of missions/projects that provided data products to the DAAC. This is a collaborative effort among the DAAC, the science teams and the missions that provide the data to the DAAC, and the plans and processes should reflect this. (Status of collecting such content should be maintained separately from this DMP and should address details for each of the datasets that is archived in the DAAC.)

7.6 Describing (documenting and preparing the information for public consumption)

7.6a – Provide pointers or references to existing technical documents related to the production of the data, for example, Algorithm Theoretical Basis Documents (ATBD), Software Interface Specifications, descriptions of products/parameters, sampling technique, instrument, data processing assumptions, etc. It is not essential to list all such documents in the DMP. A description of the types of documents maintained along with pointers to where lists of such documents are maintained will be sufficient.

7.6b - Associate data variables with quality control flag/indicator variables, if applicable. Leverage mechanisms provided in existing standards, such as CF, UMM-Var, and ISO 19157, to associate data variables with quality control flag/indicator variables, if they exist, in both data files and separate metadata records.

7.6c – Provide pointers or references to dataset user documentation that includes:

- possible error sources
- calibration method, standards, and frequency, if applicable
- dataset usage limitations (e.g., data variables intended for expert-use only, limits to temporal/geographical applicability, conditions that often result in poor data quality, intended/directed use restrictions - such as for fundamental research, education purposes, commercial purposes, or general public use).
- data gaps
- information related to a quality assessment and uncertainties of this dataset
- scientific publications

⁸ <https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/ESDIS+DOI+Process>

It is not essential to list all such documents in the DMP. A description of the types of documents maintained along with pointers to where lists of such documents are maintained will be sufficient. In the case of scientific publications, describe how the DAAC intends to capture citations and related metrics and present that information to the user.

7.6d – Describe mechanisms and plans for documentation/publications that would enable data users to access and understand data quality information. These should help address users' questions about data quality. Include informational sources outside of the DAACs where users may find additional quality information.

For example (applicable to 7.6c/d), "The [NAME] DAAC describes error sources, dataset limitation and overall quality assessment information in the User Guide for these data. Data quality information was collected by the science mission team members. Additional sources describing the data quality can be found in published research papers based on these data." List references where possible.

8. Data Discovery and Access

Describe plans and processes for ensuring that the datasets are discoverable and accessible to users. Briefly discuss ESDIS and DAAC-specific capabilities and point to applicable detailed documents.

8.1 Facilitating Discovery (*Publishing and providing access to the information*)

8.1a List applicable interfaces where a user may discover/access/filter/subset the data and data quality attributes. Include a description of how facilitating discovery follows the appropriate levels of service corresponding to the specified dataset maturity level and its compliance with NASA's open data policy as stated in section 6.1c. Note that Levels of Service are dataset-dependent, and indicate the amount of effort a DAAC spends on a dataset in various activities (such as ingest, archive, metadata development and maintenance, distribution, and user services).

Examples:

- "The datasets assigned a high Level of Service [DAAC should use the appropriate designation of LoS here] can be discovered through the DAAC's internal search interface, NASA's EarthData Search and via OPeNDAP. The dataset is free and open to the public, following NASA's Open Data Policy described in section 6.1c.
- "Datasets at beta level of maturity will be assigned a low level of service. They will not have an assigned Digital Object Identifier or landing pages. They can be discovered by contacting the user services staff at the DAAC."

8.1b Describe plans to capture and expose citation metrics for data products and how they can facilitate discovery. Describe plans to gather and expose feedback about data products from users, including targeted research users for specific data products.

8.1c Describe plans to develop and/or deploy 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. The plans should include any needed interactions with the science teams regarding proper use of their data products in the tools. For example, if the DAAC provides graphics for a variable, should the user also need to perform some quality assurance based on science team's instructions to interpret the graphics correctly?

8.1d Describe plans to engage science teams who would function as review panels and subject matter experts (SMEs) for missions/projects/PIs producing data to be delivered to the DAAC. The primary purpose is to transfer knowledge from the science team to the DAACs on the suitability and extended uses/applications of a given dataset for specific scientific disciplines and fundamental research. The DAAC should provide the names of the assigned points of contact from the DAAC and the science team/review board (if known).

For example: "The [NAME] DAAC will rely on quality information provided by the science team/review board to determine the appropriate time frame and constraints to minimize risk of misuse. This information will be provided with the data through [documentation, user announcements, etc.]. The DAAC has assigned a DAAC representative [name] to interact with the science team POCs [names], and will attend regularly scheduled science team discussions as applicable."

8.1e Describe plans for integrating a DAAC point of contact to assist the mission/project/PI as a technical data expert, as needed, in collecting known issues and other data quality attributes. This continued engagement with science teams allows the DAACs to function more directly as subject matter experts (SMEs) over the datasets they are managing.

For example: "The [NAME] DAAC has assigned a DAAC representative, [name], to interact with the mission/project/PI [names], and will attend regularly scheduled discussions as applicable."

9. User Services

Describe briefly the user services provided by the DAAC.

For example: "The [NAME] DAAC will dedicate resources to support personnel functioning as subject matter experts (SMEs) and leverage their Forum in conjunction with the EOSDIS Earthdata Support system to ensure that all user questions are categorically tracked and answered expeditiously with the proper level of technical expertise."

10. Data Rights

Generally all data are openly and freely available in accordance with NASA's Open Data Policy (see section 6.1c). Describe any exceptions here, such as International Traffic and Arms Regulations (ITAR) restrictions, Discoverability constraints, Quarantine, and Latency.

For example: The DAAC is following NASA's Open Data Policy (see section 6.1c) where applicable. The cases where data are or may not be available to the public include:

- “The [NAME] DAAC archives and maintains ITAR-controlled data which are not suitable for use outside of the mission team. These data will remain restricted.”
- “The [NAME] DAAC reserves the right to temporarily remove these data from public distribution in cases where severe or unavoidable impacts in data quality may occur. Users will be notified of such decisions and advisories that impact these data.”

11. REFERENCES

- [1] NASA 2012 NASA Space Flight Program and Project Management Requirements, https://nodis3.gsfc.nasa.gov/npg_img/N_PR_7120_005E_/N_PR_7120_005E_.pdf, NPR 7120.5E, August 14, 2012.
- [2] NASA 2008 NASA Research and Technology Program and Project Management Requirements, https://www.nasa.gov/sites/default/files/atoms/files/n_pr_7120_0008_.pdf, NPR 7120.8
- [3] NASA 2014 NASA Plans for Increasing Access to the Results of Scientific Research – Digital Scientific Data and Peer Reviewed Publications, Submitted to the Office of Science and Technology Policy on November 21, 2014, Final December 2014. https://www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf
- [4] OSTP 2013, Memorandum for the Heads of Executive Departments and Agencies, February 22, 2013, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- [5] CEOS WGISS 2016, Data Purge Alert Procedure, Version 1.0, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/WGISS_DSIG_Data%20Purge%20Alert_WP.pdf
- [6] NASA DQWG 2018, Comprehensive Data Quality Recommendations for Data Producers and Distributors, ESDS-RFC-033, <https://earthdata.nasa.gov/user-resources/standards-and-references/recommendations-from-the-data-quality-working-group>
- [7] NASA 2018, Data Management Plan Template for Data Producers, ESDS-RFC-032, <https://earthdata.nasa.gov/user-resources/standards-and-references/templates-for-nasa-data-management-plans>
- [8] Ramapriyan H. K and Moses J.F. 2013, NASA Earth Science Data Preservation Content Specification, NASA ESDIS Project, 423-SPEC-001, https://earthdata.nasa.gov/files/NASA_ESD_Preservation_Spec.pdf.

12. ACRONYMS

Abbreviation	Description
ACDD	Attribute Convention for Data Discovery
ATBD	Algorithm Theoretical Basis Document

CEOS	Committee on Earth Observation Satellites
CF	Climate Forecast
CMR	Common Metadata Repository
DAAC	Distributed Active Archive Center
DMP	Data Management Plan
DOI	Digital Object Identifier
EDOS	EOS Data and Operations System
EOSDIS	Earth Observing System Data and Information System
ESDIS	Earth Science Data and Information System
HQ	Headquarters
ISO	International Standards Organization
ITAR	International Traffic in Arms Regulations
LoS	Level of Service
MEaSURES	Making Earth Science Data Records for Use in Research Environments
NASA	National Aeronautics and Space Administration
NPR	NASA Procedural Requirements
NSIDC	National Snow and Ice Data Center
OPeNDAP	Open-source Project for a Network Data Access Protocol
OSTP	Office of Science and Technology Policy
PI	Principal Investigator
PO.DAAC	Physical Oceanography DAAC
PPBE	Planning, Programming, Budgeting, and Execution
RAC	Risk Assessment Code
SIPS	Science Investigator-led Processing System
SME	Subject Matter Expert
UMM-Var	Unified Metadata Model for Variables