Effective Date: May 2022
Expiration Date: May 2027

GSFC ESDIS CMO
May 3, 2022
Released

423-SPEC-001, Revision C
Earth Science Data and Information System Project, Code 423

NASA Earth Science Data Preservation Content Specification

Goddard Space Flight Center
Greenbelt, Maryland

National Aeronautics and
Space Administration

NASA Earth Science Data Preservation Content Specification
Signature/Approval Page

Prepared By:

Signature obtained on file — ***03/09/2022***
H. K. Ramapriyan
ESDIS Science System Consultant
NASA GSFC Code 423
Date

Signature obtained on file — ***03/09/2022***
John F. Moses
ESDIS Science System Consultant
NASA GSFC Code 423
Date

Signature obtained on file — ***03/16/2022***
Deborah Smith
ADMG Lead
NASA MSFC Code ST11
Date

Approved By:

Signature obtained on file — ***05/03/2022***
Kathleen Baynes for Kevin J. Murphy
Science Mission Directorate (SMD)
Chief Science Data Officer
NASA Earth Sciences Division
NASA HQ-DK000
Date

Signature obtained on file — ***05/03/2022***
Andrew Mitchell
ESDIS Project Manager
NASA GSFC Code 423
Date

[Electronic] Signatures available in B32 Room E148 online at: / https://ops1-cm.ems.eosdis.nasa.gov/cm2

Preface

This document is under ESDIS Project configuration control. Once this document is approved, ESDIS approved changes are handled in accordance with Class I and Class II change control requirements described in the ESDIS Configuration Management Procedures. Changes to this document shall be verified by a document change notice (DCN) and implemented by change bars or by complete revision.

Any questions should be addressed to: esdis-esmo-cmo@lists.nasa.gov

ESDIS Configuration Management Office (CMO)
NASA/GSFC
Code 423
Greenbelt, Md. 20771

Abstract

This document defines the contents of data, metadata and associated documentation to be preserved by NASA's Earth Science Division (ESD). The purpose of the document is to identify all the content items that need to be preserved to ensure their availability to support future investigations in long-term scientific research. The focus of this document is on the contents (i.e., "what" and "why") and not on the implementation or representation (i.e., "how") of the content items. Guidelines for implementation (i.e., "how") are provided in a separate document titled Preservation Content Implementation Guidance (PCIG, ESDS-RFC-042). The content items are divided into eight categories: 1. Measuring Instrument Description, 2. Instrument and Science Data Products, 3. Science Raw Data, Product and Algorithm Documentation, 4. Instrument Calibration, 5. Science Algorithm Software, 6. Science Data Product Algorithm Input, 7. Science Data Product Validation, and 8. Science Data Access and Analysis Tools. Items are described under each of these categories along with rationale for requiring their preservation. In appendices A and B, a mapping is provided between the specifications in this document and the standard ISO 19165-2:2020 titled Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products.

Keywords: Data preservation, Earth Science Data, ESDIS, OAIS

Change History Log

Revision	Effective Date	Description of Changes (Reference the CCR & CCB Approval Date)
Original	11/23/2011	CCR 423-ESDIS-4; CCB approved 11/21/2011 Pages: All
Original, CH 01	01/03/2013	CCR 423-ESDIS-28; CCB approved 11/26/2012 Pages: Title, vii, ix, 1-1, 1-2, 2-1, 3-1, 3-2, 3-5, 3-7
Revision A	01/25/2018	CCR 423-ESDIS-173, CCR approved 01/25/2018 Pages: All – Revalidation administrative extension 1 year
Revision B	01/03/2019	CCR 423-ESDIS-222; CCB approved 01/03/2018 Pages: All extend validation through September 2020
Revision C	05/03/2022	CCR 423-ESDIS-352; CCB approved 03/04/2022 Pages: Cover page through page 45 , thorough review of existing text and inclusion of statements for airborne and field investigation data. Added appendices and a reference to a Guidelines for Implementation document.

Table of Contents

1 INTRODUCTION

1.1 Scope and Background

One of NASA's strategic goals is to "Expand human knowledge through scientific discoveries" with a strategic objective to "Understand the Sun, Earth, Solar System, and Universe". NASA's Earth Science Data System (ESDS) program resides within NASA's Earth Science Division and supports the above strategic objective by providing end-to-end capabilities to deliver data and information products to users. In what follows, the word "project" is used as a general term to refer to the different types of NASA-funded activities resulting in data and derived products. Projects include on-orbit/satellite missions, suborbital/airborne and field investigations, and other data producing activities (e.g., projects under the MEaSUREs Program). The more specific terms are used only where needed to refer to the particular type of projects. The data resulting from NASA's projects are a valuable resource that needs to be preserved for the benefit of future generations. These observations are the primary record of the Earth's environment and therefore are the key to understanding how conditions in the future will compare to conditions today. In the near-term, as long as the projects' data are being used actively for scientific research, it continues to be important to provide easy access to the data and services commensurate with current information technology. For the longer term, when the focus of the research community shifts toward new projects and observations, it is essential to preserve the previous project data and associated information. This will enable a new user in the future to understand how the data were used for deriving information, knowledge, and policy recommendations and to "repeat the experiment" to ascertain the validity and possible limitations of conclusions reached in the past, and to provide confidence in long term trends that depended on data from multiple projects. It is essential for NASA to preserve all the data and associated content beyond the lives of NASA's projects to meet NASA's near-term objective of providing access to data and services for active scientific research as well as for long-term use such as reprocessing with revised algorithms to support long-term continuity with new measurements and measurement techniques. Also, NASA has to ensure that the data and associated content are preserved and available at a time in the future when permanent archive agencies will assume responsibility. To fulfill this responsibility, identification of the specific content items that need to be preserved from each of NASA's projects is essential. The purpose of this document is to specify the content items. This document focuses on the "what" (i.e., the content) and not the "how" (i.e., representation of content).

Specifications for preservation of data and associated information content complement existing archive standards. NASA and the international Consultative Committee for Space Data Systems (CCSDS) member space agencies have long recognized the importance of developing information standards for use in long-term preservation of space-related data collections. Volunteers have developed recommendations titled the "Reference Model for an Open Archival Information System (RM-OAIS)". Subsequent activities have continued to expand through a range of related interests that reach toward more practical guidance for developing agency standards. For example, they include packaging of data and metadata (XFDU), to facilitate information transfer and archiving (2008). The most recent update to the OAIS Reference Model is the Recommended Practice "Magenta Book" (2012) and supersedes the Blue Book of 2002. The CCSDS also has developed ISO 16363 that specifies requirements for certification of trustworthy digital repositories, based on the OAIS Reference Model, and ISO 16919 that describes how to audit archives for compliance with the requirements.

The OAIS Reference Model identifies an *Information Package* as "a logical container composed of optional *Content Information* and optional associated *Preservation Descriptive Information*." Content Information is defined as "a set of information that is the original target of preservation or that includes part or all of that information." Preservation Descriptive Information is defined as "the information which is necessary for adequate preservation of the Content Information and which can be categorized as *Provenance, Reference, Fixity, Context and Access Rights* information." When the datasets from projects are viewed as the Content Information, the items such as science data product documentation and processing history are part of provenance and context. However, each of these items could be considered Content Information and one can think of Preservation Descriptive Information associated with them. This specification treats such "second order" Preservation Descriptive Information as implementation details (i.e., how) and focuses on the content needed to be preserved as part of the provenance and context of the project datasets.

The Reference Model for an Open Archival Information System and related work by CCSDS does not provide guidance on the specific types of information that should be preserved with Earth science observational data. However, the Reference Model does give OAIS-compliant archives ground rules and guidance in several important areas. An OAIS-compliant archive should:

- Negotiate for and accept appropriate information from information producers
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation
- Determine, either by itself or in conjunction with other parties, which set of communities should become the "Designated Community"[1] and, therefore, should be able to understand the information provided
- Ensure that the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information
- Make the preserved information available to the Designated Community.

These guiding principles have helped in developing standard representative information requirements for Earth Science data. The first version of this specification document was developed in 2011 and approved by NASA Earth Science Division in 2013 and has been used by several EOS instrument teams for gathering and delivering preservation content to the EOSDIS DAACs. This specification, hereafter referred to as Preservation Content Specification (PCS), along with Earth Observation Preserved Data Set Content (ESA and CEOS/WGISS, 2015) has been used in the development of an international standard, ISO-19165-2:2020, titled "Geographic information -- Preservation of digital data and metadata -- Part 2: Content specifications for Earth observation data and derived digital products".

As called for by ISO-19165-2:2020, a common approach and consistency across organizations (national and international) and scientific disciplines would ensure that future long-term archives preserve necessary content, particularly since data needed for long-term science studies come from

[1] "An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time." (CCSDS, 2012).

multiple organizations and disciplinary areas. Since NASA has been using the PCS for several years, the approach for adoption of ISO 19165-2 is to provide a mapping between PCS and ISO 19165-2 (see Appendices A and B), and update sections of the PCS as needed to cover the contents called for in ISO 19165-2 that had not been previously specified in the PCS.

The content items identified in this document are based on:

- The US Global Change Research Program (USGCRP, 1998) Workshop on Global Change Science Requirements for Long-Term Archiving – October 28-30, 1998, Boulder, sponsored jointly by NASA and NOAA,
- Recent work by the ESDIS Project with several science teams whose instruments are approaching the end of their lives,
- The work by ESDIS Project staff as participants in the Earth Science Information Partner (ESIP) Federation's Data Preservation and Stewardship Cluster (ESIP, 2011), and
- ISO 19165-2.

The USGCRP workshop had participants representing a wide range of scientific disciplines. The participants developed several use cases, considering cases where:

- Existence of a data archive allowed reprocessing to produce new products for global change research,
- Existence of a data archive allowed pursuit of previously unanticipated applications,
- Lack of fully comprehensive data archives severely limited the use of data for scientific research, and
- Scientific questions and hypotheses required long-term archive services.

Based on these use cases and discussions at the workshop, several content items were identified as important for preservation.

The ESDIS Project staff has worked with many of the Earth science satellite instrument teams to identify the types of information that these teams consider important for preservation in addition to their raw data and derived products that are already in one of the EOSDIS DAACs. The discussions with these instrument teams included the items identified in the USGCRP report as a starting point in the development of PCS, and subsequently, PCS has been used for identifying the contents to be delivered to the DAACs. Also, consultation with airborne and field investigation teams has resulted in expansion of the specifications to identify appropriate preservation content for such investigations.

1.2 Organization

Preservation content artifacts are often grouped by the observing instrument/platform and all the related and necessary information that leads to the geophysical measurements (data products) derived from the instrument data. The content specification for an instrument/platform is organized into the following eight categories: 1. Measuring Instrument/Platform Description, 2. Instrument and Science Data Products, 3. Science Raw Data, Product and Algorithm Documentation, 4. Instrument Calibration, 5. Science Algorithm Software, 6. Science Data Product Algorithm Input, 7. Science Data Product Validation, and 8. Science Data Access and Analysis Tools. Each of these is described in turn in the following eight sections along with the rationale for why each of the identified items is needed. Due to differences in approaches to packaging preservation contents, it is also necessary to develop and deliver a "meta-document" that provides details of which

deliverables cover which content items. This meta-document is discussed in Section 3.9 as part of the requirement. The ESDIS Project configuration management process will be used to manage any future changes to the PCS.

The focus of this document is intentionally on the content (i.e., the "what" and "why") of items that must be preserved rather than on the method or representation (i.e., the "how"). Guidance of how to implement the specification is covered in a separate document titled Preservation Content Implementation Guidance (PCIG, ESDS-RFC-042). ISO 19165-2 addresses both "what" and "when", the latter being the stage (phase) in the project lifecycle in which the various content items should be gathered for preservation. The Appendices A and B, below, contain crosswalk tables between this document and the ISO standard. The purpose of the crosswalk tables is to show that compliance with the PCS requirements implies compliance with ISO 19165-2, as well as to guide the project phases when the preservation content should be gathered. The tables show typical items generated during a project lifecycle. The specific items, their phasing and grouping may vary from project to project. The tables should be used as a guide to make preservation as complete as possible. Judgement by the DAACs and Science Teams responsible for products will be essential to identify specific items in the project that map to the contents shown in the tables.

1.3 Related Documentation

The latest versions of all documents below should be used. The latest ESDIS Project documents can be obtained from URL: https://ops1-cm.ems.eosdis.nasa.gov. ESDIS documents have a document number starting with either 423 or 505. Other documents are available for reference in the ESDIS project library website at: https://doclib.eosdis.nasa.gov/ unless indicated otherwise.

1.3.1 Applicable Documents

The following documents are referenced within or are directly applicable, or contain policies or other directive matters that are binding upon the content of this document.

CCSDS 650.0-M-2	CCSDS, 2012: Reference Model for an Open Archival Information System (OAIS). Recommended Practice, CCSDS 650.0-M-2. Magenta Book. Issue 2. Washington, D.C.: CCSDS, June 2012. [Equivalent to ISO 14721:2003] http://public.ccsds.org/publications/archive/650x0m2.pdf
CCSDS 652.0-M-1	CCSDS, 2011: Audit and Certification of Trustworthy Digital Repositories, Recommended Practice. CCSDS 652.1-M-1. Issue 1, Washington, DC: CCSDS, September 2011 http://public.ccsds.org/publications/archive/652x0m1.pdf
CCSDS 652.1-R-1	CCSDS, 2010: Requirements for Bodies Providing Audit and Certification of Trustworthy Digital Repositories, Draft Recommended Practice. CCSDS 652.1-R-1. Issue 1, Washington, DC: CCSDS, October 2010 http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206521R1/Attachments/652x1r1.pdf
CCSDS 661.0-B-1	CCSDS, 2008: XML Formatted Data Unit (XFDU) Structure and Construction Rules. Recommendation for

	Space Data System Standards. CCSDS 661.0-B-1, Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2008.
ISO 16363:2012	ISO, 2012: Space data and information transfer systems — Audit and certification of trustworthy digital repositories, https://www.iso.org/standard/56510.html
ISO 19165-2:2020	ISO, 2020: Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products, https://www.iso.org/standard/73810.html
423-ESO-035	ESDIS, 2018: ISO 19115 Geographic Metadata Standard - Implementation Requirement and Guidance (July 2018). https://earthdata.nasa.gov/esdis/eso/standards-and-references/iso-19115
USGCRP / ESDIS01628	USGCRP, 1998: Global Change Science Requirements for Long-Term Archiving, Report of the Workshop, October 28-30, 1998, National Center for Atmospheric Research, Boulder, CO, sponsored by NASA and NOAA, through the USGCRP Program Office. [Available in ESDIS Library]
ESDS Program	"Data Processing Levels," 23 August 2019. [Online]. Available: https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels [Accessed 16 September 2020]
NASA	"NASA Strategic Plan 2018", Available: https://www.nasa.gov/sites/default/files/atoms/files/nasa_2018_strategic_plan.pdf [Accessed 23 September 2020]
ESDS-RFC-042	"Preservation Content Implementation Guidance", 2022, DOI: https://doi.org/10.5067/DOC/ESO/RFC-042

1.3.2 Reference Documents

The following documents are not binding on the content but referenced herein and, amplify or clarify the information presented in this document.

ESIP	Preservation and Stewardship Cluster of the Earth Science Information Partners Federation, 2011 https://wiki.esipfed.org/Provenance_and_Context_Content_Standard

2 CONTENT SPECIFICATIONS

2.1 Measuring Instrument/Platform Description

2.1.1 Instrument and Platform Description

- Item Description: *Documentation of instrument/sensor and platform characteristics and descriptions.* For satellite instruments this would include pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration and/or geo-location offsets). Also, for all types of instruments and platforms (e.g., spacecraft, aircraft, balloons, flux towers, and buoys), this would include items such as platform location and pointing accuracy, noise characteristics, etc. For example, components of documentation could include instrument specifications (e.g., frequencies, bandwidths, polarizations, antenna size, scan modes, etc.), vendor calibration reports, operations concepts and data acquisition timeline, spectral and radiometric calibration reports. Depending on the project documentation plans, this information may be found in a range of types of technical reports or (as in the opposite extreme), one overarching document such as a data management plan, or individual instrument journal publications.
- Rationale: Needed for users to understand how the instrument and platform operate. Documentation of measurements made before deploying instruments will help establish a baseline and help users understand changes that may have occurred over time while in operation.

2.1.2 Preflight/Pre-operational Instrument Calibration Data

- Item Description: *Numeric (digital data) files of Instrument/sensor and platform characteristics*. These should include pre-flight or pre-operational performance measurements, e.g., spectral response, instrument geometric calibration (geo-location offsets, platform location and pointing), noise characteristics, etc.
- Rationale: Measurements made before deploying instruments in space or field will help establish a baseline and help users understand changes that may have occurred over time while in operation. Also, the pre-flight calibration data could be useful for a more precise re-processing.

2.2 Instrument and Science Data Products and Metadata

2.2.1 Raw Data and Derived Products

- Item Description: *Raw data values and derived data products.* Raw data are values at full resolution (e.g., Level 0 data or Level 1a products) as directly measured by a spaceborne, airborne or field instrument. Derived products are higher level products (level 1b through 4) where calibration and geo-location transformations have been applied to generate sensor units, and/or algorithms have been applied to generate gridded geophysical parameters. A further description of processing levels can be found in ESDS Program (2019). The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the

algorithm and software that generated such earlier versions should be archived. The mapping between the software versions and data versions should be preserved.

- Rationale: Preservation of raw data (e.g., Level 0 data or Level 1a products) is required for regeneration of any higher-level products in the future in case errors are discovered or better atmospheric transmission/absorption/reflectance or scattering models become available. It is important to preserve either the means of regenerating (e.g., implementation code, software documentation, and scripts) the higher-level products or the products themselves to ensure reproducibility and verifiability of scientific results.

2.2.2 Data Product Metadata

- Item Description: *Information about data to facilitate discovery, search, access, understanding, and usage associated with each of the data products.* Metadata can be embedded in the data products or in a separate file delivered along with the data products. The metadata should include definitions of any coded values, units of measurement, references to any external standards followed, null values, etc. Metadata should follow standards according to the time of the product generation and archival (ESDIS, 2018).
- Rationale: Data cannot be located or obtained without discovery, search, and access to metadata. Data cannot be used without metadata that facilitate understanding its characteristics, quality, and usage. For example, in some production environments it becomes critical that the file-level metadata indicates which version of software was used to produce a particular data file.

2.3 Science Raw Data, Product and Algorithm Documentation

2.3.1 Instrument and Science Team

- Item Description: *Names of instrument and science team members and data product team members (development, help desk and operations), roles, organizational contact information, sponsoring agencies or organizations and comments about the products*. As responsibility changes hands over time, the names of individuals and periods during which they were responsible for various aspects of the product should be documented. Unique identifiers (e.g., ORCID) should be used for individuals.
- Rationale: It is important to know the individuals who were responsible for the instrument data products so that they are appropriately credited. Even if specific individuals are not available in the future to provide personal knowledge, their roles and responsibilities may be informative about data product quality/validity and consistency, and their publications may provide relevant insights.

2.3.2 Data Product Requirements and Designs

- Item Description: *Requirements and designs for all raw instrument/platform data and each of the derived science data products*. These may be provided either explicitly or by reference to requirements/design or data format control documents. Data product requirements and designs should include content, format, latency, accuracy, and quality. Also, any documentation such as use cases, Requests for Information (RFI), Interface Control Documents (ICD), etc. that provided inputs to the requirements development

should be preserved. If new requirements arise during a project, the new requirement documents should also be preserved.

- Rationale: Provides anticipated product outputs when the project was initiated. Useful to compare with what was actually accomplished (as recorded in validation documents). Documentation of the raw data formats and derived product formats is essential for a data product user. Detailed knowledge of the file format must be available for long-term preservation, because even file formats that are currently well understood now may become incomprehensible over time.

2.3.3 Processing and Algorithm Version History

- Item Description: *Documentation of processing history and product version history.* This should be provided for all derived data products held in the archive, indicating which versions were used when, why different versions came about, and what improvements were made from version to version. For all products held in the archive, the versions of source code used to produce the products should be available at the archive. File-level metadata should indicate which version of software was used for producing a given file. In the case of some datasets all versions of products may be maintained. In other cases, only the latest and penultimate versions may be maintained, with some samples of product files of each of the historical versions. In the case where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.
- Rationale: It is important to maintain at least the history of all the versions and the rationale for changes in order to preserve the scientific record. Traceability of inputs as well as methods that were used in generating the data product files that were used in scientific publications is essential to the scientific method.

2.3.4 Data Product Generation Algorithms

- Item Description: *Detailed discussion of processing algorithms, inputs, outputs, error budgets and limitations.* This should be provided with suggested level of detail given below:
    - Processing algorithms and their theoretical (scientific and mathematical) basis, including complete description of any sampling or mapping algorithm used in creation of the data product (e.g., such information may be contained in peer-reviewed papers and/or Algorithm Theoretical Basis Documents (ATBDs), in some cases supplemented by contextual or thematic information introducing the raw data set or derived data product)
    - Geo-location, radiometric calibration, geophysical parameters, algorithm software documentation, and high-level data flow diagrams.
    - Description of how the algorithm is numerically implemented, including possible issues with computationally intensive operations (e.g., large matrix inversions, truncation and rounding) or physical interpretation of potentially non-real mathematical results.
    - Description of all the inputs needed for the algorithm.

- Description of the output data products at a level of detail to determine if the data product has met specified product requirements.
- Description of all assumptions that have been made concerning the algorithm performance estimates and any limitations that apply to the algorithms (e.g., conditions where retrievals cannot be made or where performance may be significantly degraded.)
- Discussion of various error estimates and the error budget.

- Rationale: For any data product to be used in a scientifically valid manner, it is important to document the theoretical basis for the algorithms used to generate it and the limitations if any. The above documentation should be available and updated for each version of the data product delivered to the Data Centers so that users of a particular version know exactly how the version was generated. It is understood that for airborne and field instrument data and derived products, some of the above information will be not applicable (no algorithm is used), proprietary, unknown, or poorly documented.

2.3.5 Data Product Quality

- Item Description: *Description of the factors impacting data product quality.* Examples of causes include input data quality, issues with computationally intensive operations, large matrix inversions, truncation and rounding. Documentation of product quality assessment (methods used, assessment summaries for each version of the datasets). Description of embedded information at the file level including quality flags, product data uncertainty fields, data issue logs, etc. Relevant test reports, reviews, and appraisals. Flowed-through effects of sensor noise, calibration errors, spatial and spectral errors, and/or un-modeled or neglected geophysical phenomena on the quality of products. Description of potential future enhancements to the algorithm, the limitations they will mitigate, and other useful related information and links.
- Rationale: Users need to understand the known caveats associated with data products to ensure their proper usage. It is helpful to document any known potential improvements to algorithms that (for whatever reason) were not possible to implement for the archived data products. Information about validation and results of Calibration/Validation (Cal/Val) activities are important in characterizing data quality. They are covered in detail in section 3.7 below.

2.3.6 Data Product Application

- Item Description: *References to published articles about the use of the data and user feedback received by the science and instrument teams about the data products.* Includes reports of important scientific results, any peculiarities, or notable features observed in the products.
- Rationale: Provides additional help in the future to understand usage of the data products besides the algorithm description and source code. History of users' assessments would be useful for understanding any issues with the products. It is understood such references will need to be collected from the science teams before they disband as well as updated by the DAAC at later periods as addressed in the PCIG (ESDS-RFC-042).

2.4 Instrument Calibration

2.4.1 On-orbit, in-flight, in-the-field Calibration Methods

- Item Description: *The methods used for instrument/sensor radiometric and geometric calibration while in operation (e.g., in orbit, in flight, or in the field).* The source code used to apply the calibration algorithms. Documentation of in-line changes to calibration or to instrument or platform (spacecraft, or aircraft or field station) operations or conditions that occur throughout their operational life . In the case of field data, platforms may include fixed ground station networks, ships, buoys, radiosondes, balloons, and other types of vehicles which carry Earth observing instruments. Station location and any changes in time in location, instrumentation, controlling agency, collection protocols, surrounding land use and other factors which could influence the data record, especially for long-term data.
- Rationale: Documentation of calibration as the instrument characteristics change over time are important to be able to use data in a meaningful way, and to be able to generate long-term data time series ensuring consistency over time. Records of instrument operations history, environment and any mission anomalies are important to help understand any inconsistencies or peculiarities in data products. Source code helps users to understand how the calibrated products are generated. It may be possible to use source code (under the right circumstances) to re-implement calibration software. As stated previously in Sections 3.2.1 and 3.3.3, not all versions of data products may be preserved. However, all versions of software used for current or past public data products should be preserved and the relation between data product versions and software versions should be documented. It is understood that there may be cases where source code is proprietary and not available for archival. In those cases, such software should be identified and the reason for unavailability clearly explained along with organizational contacts and conditions under which such software (source code) could be obtained.

2.4.2 Calibration Data

- Item Description: *Instrument and platform engineering data and metadata collected during operations*. These include platform and instrument environment, events, and maneuvers: e.g., attitude and ephemeris; aircraft position; operation/acquisition logs that record data gaps; calibration look-up tables; calibration coefficients that may be delivered routinely during the life of the mission; and any significant external event data that may have impacted the observations.
- Rationale: Depending on the type of instrument and platform, many of these ancillary data types are acquired through separate data flows, distinct from the raw measurement data identified in Section 2 and are inputs necessary for on-orbit calibration and geo-location. The operations logs help users with data accountability. Users need to understand what data are available or missing from the record and why (maneuvers, anomalies, etc.)

2.5 Science Algorithm Software

- Item Description: *Science data product generation software and software documentation.* Source code used to generate data products at all levels in the science data processing system. Software release notes, including references to versions of operating systems, compilers, commercial software libraries used in the code. Versions of science data product

software should be archived for each major product release. A major product release is characterized by the appearance of peer reviewed publications where reported results are based on the product version. The mapping between the software versions and data versions should be preserved. Descriptions of data products' structure, format, range of values and special fill or error values. All information needed to verify what output data was created by a product generation job execution, including data volume, file sizes, and other information that may be needed to verify that all expected datasets were produced in the expected format. Documentation that lists the complete set of expected exceptions, and describes how they are identified, trapped, and handled. Documentation needs to identify the source for values of constants and look-up tables used in the algorithm or explain how they were calculated. The following should be included: test plans that were produced during development, including references to the artifacts; descriptions of data sets used for software verification and validation, including unit tests and system test, either explicitly or by reference to the developer's test plans; test reports or summary of the test results in sufficient detail to indicate that the data products met requirements.

- Rationale: Data product software source code and production rules provide the definitive procedural steps that document the exact implementation of the algorithm as described in ATBDs. Data product software information documents the relation between product versions and software versions. Data product software is needed when considering use of the mission collection in long multi-mission time series to understand procedural impacts relative to other instrument algorithm implementations. When examining local physical artifacts in a mission collection (spatial or temporal), product software provides a way for users to know how a particular geophysical value in the product or product metadata was derived from the combination of inputs. The data product software will enable users to know when and how extreme values or unacceptable observations were flagged and treated (e.g., not included) in a particular derived geophysical or metadata value. The data product software will help users identify the source contributions to errors and uncertainties of a particular observation. As indicated in section 3.2.1, earlier versions of software should be preserved when used to generate a version of the product that was available to the community and resulted in cornerstone findings (as advised by science community representatives). There are cases where software for some instruments may be proprietary. In such cases, the software owner must be indicated with organizational contact information.

2.6 Science Data Product Algorithm Inputs

- Item Description: *Ancillary data or other data sets used in generation or calibration of the raw data or derived data product at all levels.* Ancillary data should be stored with the data products unless it is available from another permanent archive facility. Include the name and location of the ancillary data archive facility if ancillary data will not be stored with the products. Complete information on any ancillary data or other data sets used in generation or calibration of the raw data or derived data product, either explicitly in data descriptions or by reference to appropriate publications. Information should include a full description of the input data and attributes covering all input data used by the algorithm, including primary sensor data, ancillary data, forward models (e.g., radiative transfer models, spectral line-lists, optical models, or other model that relates sensor observables to geophysical phenomena) and look-up tables. At the file level, include information on all

inputs (including ancillary or other data files, calibration files, look-up tables, ground control, climatology etc.) that were used to generate the product. At the appropriate level (file or dataset) include calibration parameters (e.g., look-up tables), precision orbit and attitude data; climatological norms, geophysical masks; First-guess fields from numerical weather or climate models; spectrum and transmittance information. Describe any important programming and procedural aspects related to implementing the algorithm into operating code.

- Rationale: The algorithm input information is needed by users investigating the data products for long multi-mission time series. Investigators need this information to understand the relative contributions of each input to an output geophysical value in the product, both at a global scale and across the life of the mission, and at local spatial (e.g., regional focus) and temporal (e.g., extreme event focus) scales. Similarly, when investigating a local physical artifact in the mission collection (e.g., regional or extreme event), the algorithm input provides a way for users to see whether the artifact is present in ancillary data such as the first-guess field, or in climate fields rather than from the instrument observations. This is especially important when investigators want to consider new improved ancillary values or ancillary geophysical relationships such as land-ocean masks or standard atmosphere profiles that could impact derived climate trends, significantly reducing error or bias in a derived product. Knowledge of all algorithm inputs is critical for assessing repeatability and usability of the experiment's results.

2.7 Science Data Product Validation

- Item Description: *Datasets and documentation; Accuracy of data products, as measured by validation testing, and compared to accuracy requirements; Description of validation process, including identification of validation data sets, measurement protocols, data collection, analysis, and accuracy reporting.* This should include a description of Cal/Val plans & status, as well as a detailed history of validation activities and validation data sets along with metadata from validation exercises. For example, ground-based target data and analysis methods should be included for laser altimeter missions. Include published data validation papers showing how well the data compare to the best-known correlative measurements.
- Rationale: Users will need to understand the procedures used for validation during the mission lifetime, as well as caveats associated with data products to ensure their proper usage. Investigators need evidence of the observed geophysical references for comparing calibrated and derived geophysical values to other long-term observational data sets. This evidence is especially important for satellite-based observations because the validation studies are often limited to comparisons with in-situ or aircraft observations from regional campaigns for finite time periods. Validation publications need to be collected from the science teams before they disband. In addition, references to additional validation activities taking place after preservation documentation has been collected will need to be routinely updated by the DAAC at later periods as addressed in the PCIG (ESDS-RFC-042).

2.8 Science Data Access and Analysis Tools

- Item Description: *Science Data access and analysis tools.* These include software source code that would facilitate use of the calibration data, ancillary data, and the raw and derived

data products at all levels. They also include software source code useful for creating programs that will read and display the calibration data, ancillary data, raw and product data as well as metadata values. Commercial tools should be identified with appropriate references. Include release notes, identify sample input, and show the corresponding output results.

- Rationale: Software tools help facilitate use of data and metadata as well as confirm documentation of the data and metadata structure. Provides an example of the data and metadata values users should expect to see from the data products.

2.9 PCS Checklist

- Item Description: *A checklist or "meta-document" that documents how each of the content specifications discussed in the above sections were satisfied.* This shall indicate whether the items were delivered by the flight project/instrument team to the assigned DAAC, to the NASA Scientific and Technical Information (STI) Program, to the ESDIS Project library, or reside in a partner archive (e.g., USGS or NOAA). This checklist should be maintained, recording any changes that may occur over time in any of the items described in sections 3.1 through 3.8 or if the archival location of some of the items are changed.
- Rationale: Different projects may have different approaches to collecting and maintaining the content items specified above. Provision of a checklist will allow for this flexibility while facilitating user access to specific items of interest. Information about points of contact will help users or DAAC personnel to obtain clarifications if needed, at least as long as the points of contact are available.

Appendix A. Mapping from PCS to ISO 19165-2

This appendix provides a mapping of the requirements described in the PCS to those in the five stages tabulated in the ISO 19165-2 standard. The content items described in the various sections of the PCS are subdivided into individual elements and assigned numbers with three decimal points (e.g., 3.1.1.1). The ISO 19165-2 requirements are numbered according to the stage to which they apply – MC, MD, MI, MO and PM. Appendix B, which provides the mapping from ISO 19165-2 to PCS, also provides the descriptions of each of the numbered requirements. The requirement numbers in the MC column should be interpreted as MC_1.1, MC_1.2, etc. shown in Appendix B. Those in the other columns should be interpreted similarly. More details about the ISO 19165-2 requirements can be found in the standard document (See References section).

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
3.1 Measuring Instrument/Platform Description						
3.1.1 Instrument and Platform Description			1.4 A-C	1.2A, 1.3, 1.4		
Instrument specifications (e.g., frequencies, bandwidths, polarizations, antenna size, scan modes)	3.1.1.1	1.2	1.1, 1.3A	1.1		
Vendor calibration reports (spectral, radiometric, geometric)	3.1.1.2			1.4		
Operations concepts	3.1.1.3	1.2	1.1	1.1, 1.2A	1.2	
Data acquisition timeline	3.1.1.4	1.1, 1.3			1.4, 1.18	
3.1.2 Preflight/Pre-operational Instrument Calibration Data			1.3B, 1.4A-C	1.4		
Spectral response	3.1.2.1			1.4		
Instrument geometric calibration (geo-location offsets)	3.1.2.2			1.4		
Noise characteristics	3.1.2.3			1.4		
3.2 Instrument and Science Data Products						

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
2.2.1 Raw Data and Derived Products	3.2.1				1.4, 1.5A-B	1.2
3.2.2 Data Product Metadata	3.2.2		1.5A	1.5.A, 1.5B	1.5A-B, 1.11	1.3
3.3 Science Raw Data, Product and Algorithm Documentation						
3.3.1 Instrument and Science Team	3.3.1	1.1				
3.3.2 Data Product Requirements and Designs	3.3.2		1.5B	1.5A, 1.6A-B		
Content	3.3.2.1			1.5A		
Format	3.3.2.2		1.5A-B	1.5A, 1.6A		
Latency	3.3.2.3	1.1		1.5A		
Accuracy	3.3.2.4	1.1		1.5A		
Quality	3.3.2.5		1.5A-B	1.5A		
3.3.3 Processing and Algorithm Version History						
Documentation of processing history and production version history (which versions were used when, why different versions came about, and what the improvements were from version to version)	3.3.3.1			1.5B	1.13	
For all products held in the archive, the versions of source code used to produce the products	3.3.3.2				1.13	
File-level metadata to indicate which version of software was used for producing a given file.	3.3.3.3				1.5A-B	

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
Where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.	3.3.3.4				1.13	
3.3.4 Data Product Generation Algorithms	3.3.4		1.5A-B			
Algorithm theoretical basis, including complete description of methods used in creation of the product (e.g., contained in peer-reviewed papers and/or Algorithm Theoretical Basis Documents (ATBDs))	3.3.4.1		1.5A	1.5B		
Algorithm software documentation	3.3.4.2			1.2C, 1.5A		
High-level data flow diagrams	3.3.4.3		1.5A			
Sampling or mapping algorithms used in creation of the product	3.3.4.4			1.5A		
geo-location, radiometric calibration, geophysical parameters	3.3.4.5			1.5A		
Description of numerical implementation, including any issues with computationally intensive operations	3.3.4.6			1.5A		
Detailed description of output data products, sufficient to determine if the product met its specified requirements.	3.3.4.7		1.5B			

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
Assumptions about algorithm performance and limitations	3.3.4.8		1.5A			
Discussion of various error estimates and the error budget.	3.3.4.9		1.3A	1.3, 1.4		
3.3.5 Data Product Quality	3.3.5				1.17	
Impact on quality due to computationally intensive operations	3.3.5.1			1.5A		
QA methods	3.3.5.2	1.2	1.6			
QA summary for each product version	3.3.5.3				1.18	
Description of QA flags	3.3.5.4			1.6A		
Description of uncertainty fields in data products	3.3.5.5			1.6A		
Test reports	3.3.5.6				1.10	
Reviews	3.3.5.7				1.10	
Appraisals	3.3.5.8				1.10	
Flowed-through effects of sensor noise, calibration errors, spatial and spectral errors, and/or un-modeled or neglected geophysical phenomena on the quality of product	3.3.5.9				1.10	
Description of potential future enhancements to the algorithm, the limitations they will mitigate	3.3.5.10				1.13	
Useful related information and links.	3.3.5.11					
3.3.6 Data Product Application						
References to published articles about the use of the data	3.3.6.1				1.19	1.4

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
User feedback about products	3.3.6.2				1.19	1.4
Reports of any peculiarities or notable features observed in the products	3.3.6.3				1.19	1.4
3.4 Instrument Calibration						
3.4.1 On-orbit, in-flight, in-the-field Calibration Methods						
Methods used for instrument/sensor radiometric and geometric calibration while in operation	3.4.1.1				1.9	
Source code used in applying the calibration algorithms	3.4.1.2				1.13	
Documentation of in-line changes to calibration	3.4.1.3				1.3, 1.18	
Documentation of changes to instrument or platform operations or conditions	3.4.1.4				1.3, 1.7	
For in situ data, station location and any changes in location, instrumentation, controlling agency, surrounding land use and other factors	3.4.1.5				1.7	
3.4.2 Calibration Data						
Platform and instrument environment	3.4.2.1				1.7	
Events and maneuvers	3.4.2.2				1.7	
Attitude and ephemeris	3.4.2.3				1.7	
Aircraft position	3.4.2.4				1.7	
Acquisition logs that record data gap	3.4.2.5				1.3	

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
Calibration look-up tables	3.4.2.6				1.8	
Calibration coefficients delivered routinely during mission	3.4.2.7				1.8	
Significant external events impacting observations.	3.4.2.8				1.8	
3.5 Science Algorithm Software	3.5					1.1
Source code used to generate products at all levels in the science data processing system.	3.5.1.1				1.13	
Software release notes, including references to versions of operating systems, compilers, commercial software libraries used in the code.	3.5.1.2				1.13	
Versions of science data product software for each major product release.	3.5.1.3				1.13	
Descriptions of data products' structure, format, range of values and special fill or error values.	3.5.1.4			1.6A		
All information needed to verify what output data was created by a run, including data volume and file sizes	3.5.1.5				1.13	
Documentation of expected exceptions, and how they are identified, trapped, and handled.	3.5.1.6				1.13	

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
Source for values of constants and look-up tables used in the algorithm, or explanation of how they were calculated.	3.5.1.7				1.13	
test plans produced during development, including references to the artifacts	3.5.1.8			1.5A, 1.7		
Descriptions of data sets used for software verification and validation	3.5.1.9				1.13	
Test reports or summary of test results in sufficient detail to show that products met requirements.	3.5.1.10			1.7	1.13	
3.6 Science Data Product Algorithm Inputs						
All ancillary data or other data sets used in generation or calibration of the data or derived product at all levels.	3.6.1.1			1.6B		
Name and location of the ancillary data archive facility if ancillary data will not be stored with the products.	3.6.1.2					1.1B
Complete information on any ancillary data or other data sets used in generation or calibration of the data set or derived product	3.6.1.3			1.6B	1.7	
Description of input data and attributes covering all input data used by the algorithm - primary	3.6.1.4				1.13	

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
sensor data, ancillary data, forward models and look-up tables.						
Information at file level on all inputs (including ancillary or other data files, calibration files, look-up tables, ground control, climatology etc.) used to generate the product.	3.6.1.5				1.7, 1.13	
At file or collection level - calibration parameters, precision orbit & attitude data; climatological norms, geophysical masks; First-guess fields from numerical weather or climate models; spectrum and transmittance information.	3.6.1.6			1.7		
3.7 Science Data Product Validation						
Accuracy of products, as measured by validation testing, and compared to accuracy requirements	3.7.1.1				1.9	
Description of validation process, including identification of validation data sets, measurement protocols, data collection, analysis and accuracy reporting.	3.7.1.2		1.4A	1.4	1.9	
Cal/Val plans & status	3.7.1.3		1.4B		1.9	

NASA Earth Science Data PCS Section #	PCS# (items in paragraphs are split and assigned numbers)	Mission Concept (MC)	Mission Definition (MD)	Mission Implementation (MI)	Mission Operations (MO)	Post-Mission (PM)
detailed history of validation activities and validation data sets along with metadata from previous validation exercises.	3.7.1.4				1.9	
published data validation papers showing how well the data compare to the best-known correlative measurements.	3.7.1.5				1.9	
3.8 Science Data Access and Analysis Tools						
Product access (reader) tools.	3.8.1.1				1.15A-B	
Source code to facilitate use of the calibration data, ancillary data, and the data products at all levels.	3.8.1.2				1.15A-B	
Source code useful for creating programs to read and display the calibration data, ancillary data and product data and metadata values.	3.8.1.3				1.15A-B	
references to applicable commercial tools	3.8.1.4				1.15A-B	
Release notes, sample inputs and corresponding output results	3.8.1.5				1.15A-B	
3.9 PCS Checklist	3.9	1.0	1.0	1.0	1.0	1.0

Appendix B. Mapping from ISO 19165-2 to PCS

This appendix provides a mapping from the requirements of preservation contents listed in the ISO 19165-2 standard to those in the PCS. The numbering of the PCS requirements corresponds to that used in Appendix A, where the items described in various numbered sections in the PCS are subdivided into individual elements. The descriptions for each of those elements can be found in Appendix A. Additional details about the requirements listed below can be found in the ISO 19165-2 standard document (see References section).

Mission Concept Stage (Phase A)

ID	Need for	Type	Identification	Description	NASA PCS section(s)
MC_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. (Note that this item will be updated at each of the subsequent stages.)	3.9
MC_1.1	ALL	Doc	Scientific/ Applications Scenario, data producer and User Communities	Defines scientific/applications scenario and expected goals. Lists Principal Investigator, designated user communities and third party actors.	3.3.1, 3.3.2.3, 3.3.2.4
MC_1.2	SAT	Doc	Mission Requirements Document	Defines scientific/applications mission and sensor requirements, processing methods, qualification methods. Includes instrument specifications (e.g., frequencies, bandwidths, polarizations, antenna size, and scan modes), data products to be produced, and operations concepts.	3.1.1.1, 3.1.1.3, 3.3.5.2
MC_1.3	ALL	Doc	Mission Operation Plan	Defines the plan for how the mission will be conducted.	3.1.1.4
MC_1.4	ALL	Doc	Mission Cost and Schedule	Defines planned cost and schedule for the mission.	None

Mission Definition Stage (Phase B)

ID	Need for	Type	Identification	Description	NASA PCS sections
MD_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. Updated version of MC_1.0	3.9
MD_1.1	ALL	Doc	Mission Requirements Specifications	Defines mission requirements, functional and resource allocation between mission measurement platform and data capture systems (e.g., ground systems for satellite operations), and operational scenario (e.g. flight plans for airplanes, un-crewed aerial vehicles, and/or drones). This may be an updated version of documentation specified in MC_1.2.	3.1.1.1, 3.1.1.3
MD_1.2	Primarily SAT	Doc	Space or aircraft to Ground segment ICDs	Defines the main systems / segments, ICDs, system latency budget estimation and data flow, including transmission of data from spacecraft to ground, transfer of recorded data on-board aircraft, handling of in situ measurement data, etc.	3.3.2
MD_1.3A	ALL	Doc	Sensor / Instrument / Platform requirements	Defines the Sensor / Instruments /Platform requirements for design (e.g. spectral bands, bandwidths, scan modes, polarizations, performance, antenna size, etc.; for photogrammetry instruments: camera/ sensors, GNSS receiver, inertial measurement unit, navigation system, etc.; requirements for platforms - satellite, aircraft, tower, buoy, etc.)	3.1.1.1, 3.3.4.9
MD_1.3B	ALL	Doc / Data Record	Sensor / Instrument processing characteristics	Characteristics for processing of acquired data, data processing model	3.1.2
MD_1.4A	ALL	Doc / Data Record	Sensor / Instrument qualification process	Qualification process for sensor, captured data, processed data.	3.1.1, 3.1.2
MD_1.4B	SAT	Doc / Data Record	Pre-launch/pre-operational calibration and characterization plan	Calibration requirements – documentation of pre-launch/pre-operational calibration methods and data from such calibration. Pre-launch calibration/pre-operational calibration includes: Optical Tests, Thermal Test, External Calibration Test, Field Of View determination	3.1.1, 3.1.2
MD_1.4C	SAT	Doc / Data Record	Ground/Ocean calibration reference and scientific base	Calibration requirements - including description of ground/ocean reference sites, accuracy, and stability of the site conditions. Data from such calibration sites.	3.1.1, 3.1.2

ID	Need for	Type	Identification	Description	NASA PCS sections
MD_1.5A	SAT and AIR; FLD if available	Doc	Processing algorithms and data format specification	Defines: Mathematical models and algorithms for mission data processing including algorithm theoretical basis; High-level data flow diagrams; Assumptions about algorithm performance and limitations; Auxiliary and ancillary data usage; Data and Product format requirements and standards; Metadata to facilitate discovery, search, access, understanding and usage associated with each of the data products.	3.2.2, 3.3.2.2, 3.3.2.5, 3.3.4
MD_1.5B	ALL	Doc	Data Product Specifications	Provides a detailed description of data products and their characteristics. It is recommended that the standard ISO 19131:2007 be followed for data product specifications. Descriptions of data products' structure, format, range of values and special fill or error values. Detailed description of output data products, sufficient to determine if products meet their specified requirements.	3.3.2.2, 3.3.2.5, 3.3.4
MD_1.6	ALL	Doc	Data Management Plan (DMP)	Preliminary DMP describing how the data and derived products will be managed during the mission	3.3
MD_1.7	ALL	Doc	Mission Cost and Schedule	Defines planned cost and schedule for the mission.	None

Mission Implementation Stage (Phases C and D)

ID	Need for	Type	Identification	Description	NASA PCS sections
MI_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. Updated version of MD_1.0	3.9
MI_1.1	ALL	Doc	Mission Design	Defines mission requirements specification and describes design as implemented.	3.1.1.1, 3.1.1.3
MI_1.2A	SAT and AIR	Doc	Detailed Space or Aircraft to Ground Segment Operations Concept and implementation	Defines the detailed operational implementation and any contingency procedure/plan needed.	3.1.1, 3.1.1.3
MI_1.2B	ALL	Doc	Updated DMP	DMP specified in MD_1.6 updated with more details and "To Be Determined (TBD)" items from the preliminary DMP filled in.	3.3
MI_1.2C	SAT	Doc	On Board Processing	Documentation of on board processing, if any.	3.3.3 (indirectly), 3.3.4.2
MI_1.3	SAT (AIR and FLD if applica-ble)	Doc	Sensor/Instrument Design and Implementation	Defines the Sensor/Instrument platform design, implementation and performance.	3.1.1, 3.3.4.9
MI_1.4	SAT, AIR	Doc / Data Records	Calibration and Validation	Independent calibration and validation campaign method, data validation activities with simulated data. Numeric (digital data) files of Instrument/sensor characteristics including pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration (geo-location offsets), and noise characteristics).	3.1.1, 3.1.1.2, 3.1.2.1, 3.1.2.2, 3.1.2.3, 3.3.4.9
MI_1.5A	SAT	Doc	Ground Processor Design, Algorithm Implementation and Supporting Information for data processing.	Defines the design and implementation of the ground data processors and the algorithm. Includes methods for geo-location, radiometric calibration, and computing geophysical parameters. Includes supporting information for data processing (e.g. ancillary, auxiliary data description & usage, etc.), and sampling or mapping algorithms used in creation of the product. Includes description of numerical implementation, including any issues with	3.3.2, 3.3.4.2, 3.3.4.4, 3.3.4.5, 3.3.4.6, 3.3.5.1, 3.5.1.8

ID	Need for	Type	Identification	Description	NASA PCS sections
				computationally intensive operations, including impact of workarounds on quality. Includes data processing software test plans and sources of test data	
MI_1.5B	ALL	Doc	Technical Notes Scientific Papers	Algorithm description and software validation for all software used on ground and on board. Metadata and naming conventions	3.2.2, 3.3.3.1, 3.5
MI_1.6A	ALL	Doc	Data Format Specifications	Contains information that will allow the user to read and use the data. Includes data format standard(s) used.	3.3.2.2, 3.3.5.4, 3.3.5.5, 3.5.1.4
MI_1.6B	ALL	Doc / Data Records	Supporting Information for processing	Defines and identifies ancillary and auxiliary data and how they are to be used in processing. Documentation about and data records for all ancillary data or other data sets used in generation or calibration of the data or derived product at all levels.	3.3.2, 3.6.1.1, 3.6.1.3
MI_1.7	ALL	Doc / Data Records	Qualification Process	Detailed qualification methods and data	3.5.1.8, 3.5.1.10

Mission Operation Stage (Phase E)

ID	Need for	Type	Identification	Description	NASA PCS Sections
MO_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. Updated version of MI_1.0	3.9
MO_1.1	SAT	Doc	Mission data access and service requirements document and User Handbook	Defines the data archival and processing/ reprocessing strategy, the data accessible to users and the service requirements & performance during the operations stage.	None
MO_1.2	SAT	Doc	Sensor Ground Segment Operations Plan	Describes the actual implementation of the end-to-end mission operations.	None
MO_1.3	ALL	Doc/ Data Records	Mission Operations Acquisition Plans and Reports	Describes acquisition plans and reports for data from the mission sensor(s). Includes mission logs accounting for data gaps, maneuvers, etc.	3.4.2.5
MO_1.4	ALL	Data Records	Raw/Level 0	Raw or Level 0 data from the sensor or instrument data packets	3.2.1
MO_1.5A	ALL	Data Records	Level 1	Processed science data L1 products. The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the algorithm and software that generated earlier versions should be archived. Metadata at the dataset (file) level should indicate which version of software was used for producing a given dataset (file).	3.2.1, 3.3.3.3
MO_1.5B	ALL	Data Records	Level 2 +	Processed science data L2 products and higher, including Analysis Ready Data. The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the algorithm and software that generated earlier versions should be archived. Metadata at the dataset (file) level should indicate which version of software was used for producing a given dataset (file). Metadata at the dataset (file) level, or at	3.2.1, 3.3.3.3

ID	Need for	Type	Identification	Description	NASA PCS Sections
				data set series (collection) level as appropriate, should include information about calibration parameters, precision orbit & attitude data; climatological norms, geophysical masks; first-guess fields from numerical weather or climate models; spectrum and transmittance information; ; detailed site descriptions and field methodology should be documented for in-situ sampling.	
MO_1.6	ALL	Data Records	Browse Images	Browse Digital Catalogue and images, when generated.	None
MO_1.7	SAT, AIR (FLD if needed)	Data Records	Ancillary Data	Attitude, Ephemeris, Navigation parameters, Observation counters, Orbital State Vectors, Times, Sun Position, Temperatures Sensor/CCD/Amplifiers noises, Earth Relative position, Azimuth, satellite or aircraft maneuvers, instrument parameters (e.g. optical response), aircraft position, locations of in situ instruments	3.4.1.4, 3.4.1.5, 3.4.2, 3.6.1.3, 3.6.1.5
MO_1.8	SAT (AIR/ FLD if available)	Data Records	Auxiliary Data	Band/Multispectral/ Band-by-band parameters for algorithms, Non-linearity correction factors, Error/Failure/Gap correction factors, Calibration curve/Factors or look-up tables, Scaling correction factors, Atmospheric correction factors, geometry correction factors, drift factor, albedo parameters, instrument modes, incident angle, absolute calibration constants, solar radiance, moon temperature brightness, local seasonal variances, weather forecast/actual, wind, altimetry/geoid model DEM, any significant external events impacting observations.	3.4.2, 3.6
MO_1.9	ALL	Doc/ Data Records	Calibration and validation data	Description of validation process, including identification of validation data sets, measurement protocols, data collection, analysis and accuracy reporting. Cal/Val (CALibration / VALidation) data acquired during mission operations (optical/radiometric stability, Instrument availability, Internal calibration, Optic pointing pattern, etc.) Detailed history of validation activities and validation data sets along with metadata from previous validation	3.4.1.1, 3.4.1.3, 3.4.2, 3.6, 3.7

ID	Need for	Type	Identification	Description	NASA PCS Sections
				exercises. Published data validation papers showing how well the data compare to the best known correlative measurements. Accuracy of products, as measured by validation testing, and compared to accuracy requirements.	
MO_1.10	ALL	Doc/ Data Records	Quality Parameters	Quality assessment of instrument, raw data and products Flowed-through effects of sensor noise, calibration errors, spatial and spectral errors, and/or un-modelled or neglected geophysical phenomena on the quality of product	3.3.5.6, 3.3.5.7, 3.3.5.8, 3.3.5.9
MO_1.11	ALL	Doc/ Data Records	Metadata	Metadata Digital Inventory. Information about data to facilitate discovery, search, access and use associated with each of the data products.	3.2.2
MO_1.12	SAT	SW Code	Level 0 consolidation	Software used for generating Level 0 data from raw data.	None
MO_1.13	SAT, AIR	Doc/ SW Code	Data Processing Software	Instrument processing algorithms, context and source codes, testing context. Description of lineage – i.e., input data and attributes covering all input data used by the algorithm - primary sensor data, ancillary data, forward models and look-up tables. Lineage information at file level – i.e., all inputs (including ancillary or other data files, calibration files, look-up tables, ground control, climatology etc.) used to generate the product. All information needed to verify what output data was created by a run, including data volume and file sizes. Documentation of expected exceptions, and how they are identified, trapped, and handled. Source for values of constants and look-up tables used in the algorithm, or explanation of how they were calculated. Documentation of processing history and production version history (which versions were used when, why different versions came about, and what the improvements and changes were from version to version). Descriptions of data sets used for software verification and validation. Test reports or summary of test results in sufficient detail to show that products meet requirements. Software release notes, including references to versions of operating	3.3.3.1, 3.3.3.2, 3.3.3.4, 3.3.5.10, 3.4.1.1, 3.4.1.2, 3.5

ID	Need for	Type	Identification	Description	NASA PCS Sections
				systems, compilers, commercial or other software libraries used in the code.	
MO_1.13 (Continued)				Description of potential future enhancements to the algorithm, the limitations they will mitigate. For all products held in the archive, the versions of source code used to produce the products. Where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.	3.3.3.1, 3.3.3.2, 3.3.3.4, 3.3.5.10, 3.4.1.1, 3.4.1.2, 3.5
MO_1.14	SAT	SW Code	Quality Control Software	Software used for quality assessment and developing quality indicators and/or quality flags.	3.3.5
MO_1.15A	SAT	SW Code	Science Data Tools	Product access (reader) and analysis tools. Source code to facilitate use of the calibration data, ancillary data and the data products at all levels. Source code useful for creating programs to read and display the calibration data, ancillary data and product data and metadata values. References to applicable dependency tools and libraries, and version numbers. Release notes, sample inputs and corresponding output results.	3.8
MO_1.15B	SAT	SW Code	Visualization Tools	Processing and visualizing tools	3.8 (not explicitly)
MO_1.16	SAT	SW Code	Value-Added Software	Software used for developing value-added products from the basic products (see MO_1.5A and MO_1.5B).	3.8 (not explicitly)
MO_1.17	SAT	Doc	Product qualification and quality assurance monitoring reports	Defines the product qualification process outputs. Includes test reports, and results of reviews and appraisals.	3.3.5, 3.5
MO_1.18	ALL	Doc	Sensor/ Instrument evolution and history records	Describes any instrument event that might affect data quality (e.g. upgrading, downgrading, look-up tables). It includes also known-errors and limits of sensors/instruments. This is essential, especially for airborne and in-situ observations where sensors are updated or replaced during a mission. All calibration and quality assessment results/processes/reports should be explicitly connected to each instance of sensors/instruments.	3.3.5.3, 3.4.1.3, 3.4.1,4

ID	Need for	Type	Identification	Description	NASA PCS Sections
MO_1.19	ALL	Doc	Referred publications and papers	Referred publications, articles and technical notes clearly referencing the used datasets. References to published articles about the use of the data User feedback about products	3.3.4, 3.3.6, 3.7
MO_1.20	ALL	Doc	Tandem and/or combined campaigns, comparisons	Data and reports	3.3.5
MO_1.21	ALL	Doc/ Data Records	Cross-campaign, cross-comparisons and cross-calibration activities documentation and Data	Describes the cross-campaign scenario and operational context. Also describes any cross-calibration activities.	3.4 (cross-calibration not explicit)
MO_1.22	ALL	Doc	Data Access Policy	Describes the data access policy for mission in the operational stage.	None

Post-Mission Stage (Phase F)

ID	Need for	Type	Identification	Description	NASA PCS Sections
PM_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. Updated version of MO_1.0	3.9
PM_1.1A	SAT	Doc	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Processing	Description of final processing and/or calibration changes including provenance and context.	3.5
PM_1.1B	SAT	Doc/ Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Ancillary, Auxiliary	Updated Ancillary, Auxiliary data and their description. Name and location of the ancillary/auxiliary data archive facility if ancillary/auxiliary data will not be stored with the products.	3.4.2, 3.6.1.2
PM_1.1C	SAT	Doc/ Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. QA	Quality information updated as part of reprocessing	3.3.5
PM_1.2	ALL	Data Records (Reprocessed data set)	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. L0, L1, L2+	Reprocessed data & products The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the algorithm and software that generated such earlier versions should be archived.	3.2.1
PM_1.3	ALL	Data	Data consolidation & reprocessing strategy, implementation plans and consolidated/ reprocessed data. Metadata	Metadata Inventory	3.2.2
PM_1.4	ALL	Doc	Referred publications and papers	Referred publications, articles and technical notes clearly referencing the used datasets.	3.3.4, 3.3.6, 3.7

ID	Need for	Type	Identification	Description	NASA PCS Sections
PM_1.5	SAT	Doc	Historical Data Access Policy	Describes the data access policy for the historical mission in the Preservation stage.	None
PM_1.6	SAT	Doc	Historical Mission User Handbook	Describes the consolidated end-to-end mission description, data formats, operational scenarios, and all information necessary for future data use. It also includes the appraisal of the mission datasets (i.e., their value).	None

Appendix C. Abbreviations and Acronyms

ADMG	Airborne Data Management Group
AIR	Airborne Imaging Radar
ATBD	Algorithm Theoretical Basis Document
Cal/Val	Calibration/Validation
CCB	Configuration Control Board
CCD	Charged-Coupled Device
CCR	Configuration Change Request
CCSDS	Consultative Committee for Space Data Systems
CEO	Committee on Earth Observation Satellites
CMO	Configuration Management Office
DAAC	Distributed Active Archive Center
DCN	Document Change Notice
DEM	Digital Elevation Model
DMP	Data Management Plan
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
ESA	European Space Agency
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System
ESDS	Earth Science Data System
ESIP	Federation of Earth Science Information Partners
GNSS	Global Navigation Satellite System
GSFC	Goddard Space Flight Center
ICD	Interface Control Document
IMPACT	Interagency Implementation and Advanced Concepts Team
ISO	International Organization for Standardization
L0-L2	Level 0-2
MC	Mission Concept
MD	Mission Definition
MEaSUREs	Making Earth System Data Records (ESDR) for Use in Research Environments
MI	Mission Implementation
MO	Mission Operations
MSFC	Marshall Space Flight Center
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OAIS	Open Archival Information System
PCIG	Preservation Content Implementation Guidance
PCS	Preservation Content Specification
PM	Post-Mission
QA	Quality Assurance
RFI	Requests for Information
RM	Reference Model
RM-OAIS	Reference Model for an Open Archival Information System

STI	Scientific and Technical Information
SW	Shortwave radiation
TBD	To Be Determined
USGCRP	United States Global Change Research Program
USGS	United States Geological Survey
WGISS	Working Group on Information Systems and Services
XFDU	XML Formatted Data Unit
XML	eXtensible Markup Language