ISO/TC 211
Geographic information/Geomatics

Email of secretary: therese.andren@sis.se
Secretariat: SIS (Sweden)

ISO/NP 19165-2 (Earth observation data and derived digital products)

Document type:	NP ballot
Date of document:	2017-05-12
Expected action:	VOTE
Action due date:	2017-08-07
No. of pages:	65
Background:	The ballot of ISO/NP 19165-2 (N 4518) is launched in ISO eBalloting Portal from 2017-05-15 to 2017-08-07. Please note that official balloting is for internal balloters only. Experts and delegates have to reply to the relevant national standards body.
Committee URL:	http://isotc.iso.org/livelink/livelink/open/tc211

Form 4: New Work Item Proposal

Circulation date: 2017-05-12 Closing date for voting: 2017-08-07	Reference number: ISO/NP 19165-2 (to be given by Central Secretariat) ISO/TC 211 **N 4518**
Proposer (e.g. ISO member body or A liaison organization) ANSI	
Secretariat SIS	

A proposal for a new work item within the scope of an existing committee shall be submitted to the secretariat of that committee with a copy to the Central Secretariat and, in the case of a subcommittee, a copy to the secretariat of the parent technical committee. Proposals not within the scope of an existing committee shall be submitted to the secretariat of the ISO Technical Management Board.

The proposer of a new work item may be a member body of ISO, the secretariat itself, another technical committee or subcommittee, an organization in liaison, the Technical Management Board or one of the advisory groups, or the Secretary-General.

The proposal will be circulated to the P-members of the technical committee or subcommittee for voting, and to the O-members for information.

☒ The proposer has considered the guidance given in the Annex C during the preparation of the NWIP.

Proposal (to be completed by the proposer)

Title of the proposed deliverable.

English title:

Geographic information -- Preservation of digital data and metadata -- Part 2: Content specifications for earth observation data and derived digital products

French title:

Titre manque

(In the case of an amendment, revision or a new part of an existing document, show the reference number and current title)

Scope of the proposed deliverable.

This standard is being proposed as an extension to ISO 19165 Geographic information – Preservation of Digital Data and Metadata, which at this time has gone through an editing cycle as a Committee Draft. This extension, proposed as ISO 19165-2, will provide a description of the contents that organizations must preserve to ensure that the data and derived digital products resulting from Earth observing missions (remote sensing with airborne and spaceborne instruments, and in situ measurements such as those from field campaigns) remain usable and understandable in order to support long-term future scientific research. This document will cover what content items must be preserved, but is not concerned about how they are to be preserved, what their formats should be, etc.

This part of ISO 19165 extends the standard for the long-term preservation of digital geospatial data to provide details about the content describing the provenance and context specific to data from Earth observing missions. It also applies the ISO standards ISO 19115-1 “Metadata – Part 1: Fundamentals”, ISO 19115-2 “Metadata - Part 2: Extensions for imagery and gridded data”, and ISO 19157:2013 “Geographic information -- Data quality” where appropriate.

This standard takes advantage of the work performed by the Data Preservation and Stewardship Committee of the U.S. Federation of Earth Science Information Partners (ESIP), NASA, ESA, and CEOS Working Group on Information Systems and Services (WGISS). The documents from these groups are integrated, along with the ISO standards mentioned above to provide specific content items to be preserved from Earth observing missions for the benefit of users. It is expected that if the content items specified by this standard are preserved, users will have sufficient information to be able to understand, reuse, and ideally regenerate data products without the assistance of the original teams that were responsible for their initial generation.

Purpose and justification of the proposal*

See Annex A “PCS Justification – 20161223”

Consider the following: Is there a verified market need for the proposal? What problem does this standard solve? What value will the document bring to end-users? See Annex C of the ISO/IEC Directives part 1 for more information. See the following guidance on justification statements on ISO Connect:
https://connect.iso.org/pages/viewpage.action?pageId=27590861

Preparatory work (at a minimum an outline should be included with the proposal)

☐ A draft is attached ☐ An outline is attached ☒ An existing document to serve as initial basis

The proposer or the proposer's organization is prepared to undertake the preparatory work required:

☒ Yes ☐ No

If a draft is attached to this proposal:

Please select from one of the following options (note that if no option is selected, the default will be the first option):

☒ Draft document will be registered as new project in the committee's work programme (stage 20.00)

☐ Draft document can be registered as a Working Draft (WD – stage 20.20)

☐ Draft document can be registered as a Committee Draft (CD – stage 30.00)

☐ Draft document can be registered as a Draft International Standard (DIS – stage 40.00)

Is this a Management Systems Standard (MSS)?

☐ Yes ☒ No

NOTE: if Yes, the NWIP along with the Justification study (see Annex SL of the Consolidated ISO Supplement) must be sent to the MSS Task Force secretariat (tmb@iso.org) for approval before the NWIP ballot can be launched.

Indication(s) of the preferred type to be produced under the proposal.

☒ International Standard ☐ Technical Specification

☐ Publicly Available Specification ☐ Technical Report

Proposed development track

☐ 1 (24 months) ☒ 2 (36 months - default) ☐ 3 (48 months)

Note: Good project management is essential to meeting deadlines. A committee may be granted only one extension of up to 9 months for the total project duration (to be approved by the ISO/TMB).

Known patented items (see ISO/IEC Directives, Part 1 for important guidance)

☐ Yes ☒ No

If "Yes", provide full information as annex

Co-ordination of work: To the best of your knowledge, has this or a similar proposal been submitted to another standards development organization?

☐ Yes ☒ No

If “Yes”, please specify which one(s):

A statement from the proposer as to how the proposed work may relate to or impact on existing work, especially existing ISO and IEC deliverables.
The proposer should explain how the work differs from apparently similar work, or explain how duplication and conflict will be minimized.

There are several related ISO standards that provide the basis for the proposed new work item. However, they do not directly address the specification of preservation content for Earth observation data to be addressed by this NWIP. The relevant ISO standards are discussed below.

ISO 14721:2012, 16363 and 16919: Space agencies that are members of the international Consultative Committee for Space Data Systems (CCSDS) 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 Open Archival Information System (RM-OAIS). Subsequent activities continue to expand through a range of related interests that reach toward more practical guidance for developing agency standards. They include provider-archive interchange recommendations (2004) and packaging of data and metadata (XFDU), to facilitate information transfer and archiving (2008). The most recent update to the OAIS Reference Model is the publication ISO 14721:2012 “Space data and information transfer systems – Open archival information

A statement from the proposer as to how the proposed work may relate to or impact on existing work, especially existing ISO and IEC deliverables.
The proposer should explain how the work differs from apparently similar work, or explain how duplication and conflict will be minimized.

system (OAIS) – Reference model". The OAIS Reference Model provides a conceptual archival framework. 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.

ISO 19115: ISO 19115 provides a metadata model for describing geographic information and services, and ISO 19115-2 augments ISO 19115 with additional structure to describe more extensively the derivation of geographic imagery and gridded data. It provides the structure needed to represent properties of instrument acquiring data, instrument geometry, production processes, etc. The structure provided by ISO 19115-2 is useful for representing the preservation content intended to be specified with this NWIP.

ISO 19165: ISO 19165, a standard for "Geographic Information - Preservation of digital data and metadata" has been under development since 2014 and its Committee Draft was edited at the TC 211 meeting in Redlands, CA during November 29-20, 2016. ISO 19165 considers geographic information preservation in general and is related to the present NWIP. As stated in ISO 19165, "In preserving data, future users need to understand what they are working with (context information) and how the data were created (provenance information). Because most Earth science data involve complex physics and mathematics, the metadata shall include sufficient documentation (or pointers thereto) that provide the derivation of the algorithms used to generate the dataset. Likewise, the metadata shall include pointers to calibration data and ancillary data that were needed to produce the dataset. The specific content items needed to preserve the full provenance and context of the data and associated metadata depend on the needs of the designated community and types of datasets (e.g., maps, remotely sensed data from satellites and airborne instruments, physical samples). Follow-up parts to this standard may be developed detailing content items appropriate to individual disciplines." The present NWIP is for developing more detailed specifications for remotely sensed Earth science data and derived digital products, and will be an extension (say ISO 19165-2) to ISO 19165.

More details are provided in Annex A "PCS Justification – 20161223"

A listing of relevant existing documents at the international, regional and national levels.

ISO 14721: 2012, Space data and information transfer systems – Open archival information system (OAIS) – Reference model
ISO 19115-2, Geographic information - Metadata - Part 2: Extensions for imagery and gridded data
NASA (2011, revised 2013) Earth Science Preservation Content Specification, https://earthdata.nasa.gov/sites/default/files/field/document/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf
ESA (2012), Long Term Data Preservation, Earth Observation Preserved Data Set Content (LTDP/PDSC), http://earth.esa.int/gscb/ltdp/LTDP_PDSC_4.0.pdf
CEOS Earth Observation Preserved Data Set Content (PDSC), http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf
ISO 19165 (Committee Draft): Geographic Information - Preservation of digital data and metadata

Please fill out the relevant parts of the table below to identify relevant affected stakeholder categories and how they will each benefit from or be impacted by the proposed deliverable(s).

	Benefits/impacts	Examples of organizations / companies to be contacted
Industry and commerce large industry		
Industry and commerce SMEs		

Government	Helps with planning for and management of data from Earth observing missions to ensure consistent preservation of data and their understandability for future users.	Canada Centre for Mapping and Earth Observation (CCMEO), CEOS, CNES, DLR, ISRO, JAXA, NASA, National Remote Sensing Center of China (NRSCC), NOAA, Russian Federal Space Agency (ROSKOSMOS), USGS, etc.
Consumers	Users of Earth observation data products will have sufficient information to be able to understand, reuse, and, ideally, regenerate data products without the assistance of the original teams that were responsible for their initial generation.	
Labour		
Academic and research bodies		
Standards application businesses		
Non-governmental organizations		
Other (please specify)		

Liaisons:

A listing of relevant external international organizations or internal parties (other ISO and/or IEC committees) to be engaged as liaisons in the development of the deliverable(s).

ISPRS, CEOS, IEEE GRSS

Joint/parallel work:

Possible joint/parallel work with:

- ☐ IEC (please specify committee ID)
- ☐ CEN (please specify committee ID)
- ☐ Other (please specify)

A listing of relevant countries which are not already P-members of the committee.

Note: The committee secretary shall distribute this NWIP to the countries listed above to see if they wish to participate in this work

Proposed Project Leader (name and e-mail address)

Hampapuram Ramapriyan
hampapuram.ramapriyan@ssaihq.com

Name of the Proposer
(include contact information)

Hampapuram Ramapriyan
hampapuram.ramapriyan@ssaihq.com

This proposal will be developed by:

☒ An existing Working Group: ISO/TC 211/WG 7

☐ A new Working Group:

(Note: establishment of a new WG must be approved by committee resolution)

☐ The TC/SC directly

☐ To be determined:

Supplementary information relating to the proposal

☒ This proposal relates to a new ISO document

☐ This proposal relates to the adoption as an active project of an item currently registered as a Preliminary Work Item

☐ This proposal relates to the re-establishment of a cancelled project as an active project

Other:

☒ Annex(es) are included with this proposal (give details)

Annex A PCS Justification – 20161223
Annex B CEOS Earth Observation Preserved Data Set Content (PDSC)
Annex C NASA Earth Science Preservation Content Specification (PCS)

Additional information/question(s)

Annexes B and C listed above will be the starting point for the development of a Working Draft for this NWIP.

Justification for New Work Item Proposal to develop "Preservation Content Specifications for Earth Observation Data and Derived Digital Products"

Information from Earth observing missions (remote sensing with airborne and spaceborne instruments, and in situ measurements such as those from field campaigns) is proliferating in the world. Many agencies across the globe are generating important datasets by collecting measurements from instruments on board aircraft and spacecraft, globally and constantly. The data resulting from such measurements 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. Earth science observational data, derived products and models are used to answer key questions such as "How is the Global Earth System Changing?", "What are the sources of change in the Earth systems and what are their magnitudes and trends?", "How will Earth system change in the future?", and "How can Earth system science improve mitigation of and adaptation to global change?" In the near-term, as long as the missions' 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 missions and observations, it is essential to preserve the previous mission 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 missions.

Organizations that collect, process, and utilize Earth observation data today have a responsibility to ensure that the data and associated content continue to be preserved by them or are gathered and handed off to other organizations for preservation for the benefit of future generations. In order to ensure preservation of complete content necessary for understanding and reusing the data and derived digital products from today's missions, it is necessary to develop a specification of such preservation content. While there are existing standards that address archival and preservation in general, there are no existing international standards or specifications today to address what content should be preserved.

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 information content complement existing archive standards. Space agencies that are members of the international Consultative Committee for Space Data Systems (CCSDS) 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 Open Archival Information System (RM-OAIS). Subsequent activities continue to expand through a range of related interests that reach toward more practical guidance for developing agency standards. They include provider-archive interchange recommendations (2004) and packaging of data and metadata (XFDU), to facilitate information transfer and archiving (2008). The most recent update to the OAIS Reference Model is the publication ISO 14721:2012 "Space data and information

transfer systems – Open archival information system (OAIS) – Reference model". The OAIS Reference Model provides a conceptual framework for archival. 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 missions 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 preserved as part of the provenance and context of the mission datasets.

The PREMIS international standard for metadata addresses preservation of digital objects and ensuring their long-term usability. It provides a detailed data dictionary applicable in general to digital object preservation. However, it does not provide specifics of the content types that need to be preserved for Earth science datasets.

ISO 19115 provides a metadata model for describing geographic information and services, and ISO 19115-2 [2] augments ISO 19115 with additional structure to describe more extensively the derivation of geographic imagery and gridded data. It provides the structure needed to represent properties of instrument acquiring data, instrument geometry, production processes, etc. The structure provided by ISO 19115-2 is useful for representing the preservation content intended to be specified with this NWIP.

ISO 19165, a standard for "Geographic Information - Preservation of digital data and metadata" has been under development since 2014 and its Committee Draft was edited at the TC 211 meeting in Redlands, CA during November 29-20, 2016. ISO 19165 considers geographic information preservation in general and is related to the present NWIP. As stated in ISO 19165, "In preserving data, future users need to understand what they are working with (context information) and how the data were created (provenance information). Because most Earth science data involve complex physics and mathematics, the metadata shall include sufficient documentation (or pointers thereto) that provide the derivation of the algorithms used to generate the dataset. Likewise, the metadata shall include pointers to calibration data and ancillary data that were needed to produce the dataset. The specific content items needed to preserve the full provenance and context of the data and associated metadata depend on the needs of the designated community and types of datasets (e.g., maps, remotely sensed data from satellites and airborne instruments, physical samples). *Follow-up parts to this standard may be developed detailing content items appropriate to individual disciplines*." The present NWIP is for developing more detailed specifications for remotely sensed Earth science data and derived digital products, and will be an extension (say ISO 19165-2) to ISO 19165.

There are several other efforts addressing preservation of digital data and metadata. Many of these are listed in the references section below [1-9]. They provide useful information for preservation, metadata to be considered, representation, formats, procedures, appraisal and selection, architectures, etc.

However, none of them addresses the specific contents that must accompany remotely sensed Earth science data and derived digital products to ensure their long-term understandability and usability.

Given the lack of a standard for such preservation content, the Data Preservation and Stewardship Committee of the U.S. Federation of Earth Science Information Partners (ESIP) drafted an "emerging" Provenance and Context Content Standard (PCCS) [10] in January 2011. Representatives from several U.S. agencies are involved in this committee, including the National Oceanic and Atmospheric Administration (NOAA) and the United States Geological Survey (USGS). The PCCS was developed based on:

1. a report of a workshop on long-term archiving conducted by the US Global Research Program in 1998 [11];
2. NASA's experience with some of its space-borne remote sensing instrument teams; and
3. Inputs from NOAA regarding their documentation requirements and practices.

In late 2011, NASA developed a document titled "NASA Earth Science Data Preservation Content Specifications" [12]. NASA has been using this as a requirement for new Earth Science missions and as a checklist for missions that are completed or currently in operation (given that they had not been provided with the specifications as a requirement). In parallel with these efforts in the U.S., the European Space Agency has been actively pursuing Long-Term Data Preservation (LTDP) since 2008 in coordination with other space agencies in Europe [i.e. Italian Space Agency (ASI), French Space Agency (CNES), German Space Agency (DLR), Canadian Space Agency (CSA), European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), United Kingdom Space Agency (UKSA)] in the frame of the Ground Segment Coordination Body (GSCB) and has generated several pertinent Best Practices (e.g., [13]) addressing both the preservation of data and associated knowledge, but also identifying the EO mission data set assets content (i.e. data records and associated knowledge) that should be preserved to ensure long-term usability and exploitation of Earth Science data. These Best Practices have been subsequently reviewed and consolidated in the frame of the Committee on Earth Observation Satellites (CEOS) including feedback and lessons learned from other Space Agencies worldwide [e.g., 14-16].

Given the progress made by the US and European agencies in identifying preservation content, we believe that we have a good starting point for developing an ISO TC 211 specification titled as proposed here.

References:

1. OAIS (2012): Space data and information transfer systems – Open archival information system (OAIS) – Reference model, ISO 14721:2012, http://public.ccsds.org/publications/archive/650x0m2.pdf
2. ISO 19115-2, Geographic information - Metadata - Part 2: Extensions for imagery and gridded data, http://www.iso.org/iso/catalogue_detail.htm?csnumber=39229.
3. PREMIS Data Dictionary for Preservation Metadata, July 2012, http://www.loc.gov/standards/premis.
4. Sarah Higgins (2008): The DCC Curation Lifecycle Model, in: International Journal of Digital Curation, vol. 3 http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48

5. Sabine Schrimpf, Christian Keitel, nestor Guideline for Preservation Planning – a Process Model, URN:NBN:NL:UI:10-1-116062, (2014) LIBER Quarterly, volume 23, issue 3, pp. 201 – 213
6. Archiving of Geodata (2010). A joint preliminary study by swisstopo and the Swiss Federal Archives http://www.swisstopo.admin.ch/internet/swisstopo/en/home/topics/geodata/geoarchive.parsysrelated1.59693.downloadList.93958.DownloadFile.tmp/preliminarystudyarchivingofgeodata.pdf
7. G. McGarva, S. Morris, G. Janée (2009) DPC Technology Watch Report on preserving geospatial data http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee
8. G. Janée, J. Sweetkind-Singer, T. Moore (2009): Final Report of the National Geospatial Digital Archive (NGDA) and Federated Archive Cyberinfrastructure Testbed (FACIT) Projects, http://www.ngda.org/docs/ngda-final-report.pdf
9. Steve Morris (2010): Appraisal and Selection of Geospatial Data White Paper, Prepared for Library of Congress http://www.digitalpreservation.gov/meetings/documents/othermeetings/AppraisalSelection_whitepaper_final.pdf
10. ESIP Federation Provenance and Context Content Standard (PCCS): ESIP Provenance and Context Content Standard (wiki), 2011
11. USGCRP, http://www.globalchange.gov/browse/reports/global-change-science-requirements-long-term-archiving, doi:10.7930/J0CZ353N
12. NASA (2011) Earth Science Preservation Content Specification (PCS): NASA Earth Science Data Preservation Content Specification (PDF)
13. https://earth.esa.int/web/gscb/ltdp
14. CEOS Earth Observation Preserved Data Set Content (PDSC), http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf
15. CEOS Earth Observation Preservation Guidelines, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Data%20Preservation%20Guidelines_v1.1.pdf
16. CEOS Earth Observation Preservation Workflow, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf

Long Term Preservation of Earth Observation Space Data

Earth Observation Preserved Data Set Content

CEOS-WGISS

Data Stewardship Interest Group

Doc. Ref.: CEOS/WGISS/DSIG/EOPDSC

Date: September, 2015

Issue: Version 1.0

Document Status Sheet

Issue	Date	Comments	Editor
1.0	15 September 2015	New document evolved from EO Preserved Data Set Content, Issue 4.0	M. Albani, R.Leone, I. Maggio, R. Cosac

Table of Contents

1. INTRODUCTION **1**
1.1 Intended Audience **1**
1.2 Background **1**
1.3 Document Scope **1**
1.4 Applicable and Reference Documents **2**

2. HOW TO USE THIS EO PDSC **3**
2.1 "What" Dimension **3**
2.2 "When" Dimension **3**
2.3 "Quality" Dimension **3**
2.4 "Preservation metadata" Dimension **4**
2.5 "How" Dimension: Tailoring the PDSC **4**

3. PRESERVED DATA SET CONTENT SPECIFICATION **6**
3.1 Data Records **6**
3.2 Tools **7**
3.3 Information **7**

4. PRESERVED DATA SET CONTENT FOR EARTH OBSERVATION MISSIONS **9**
4.1 Mission Concept Stage (MC) **10**
4.2 Mission Definition Stage (MD) **11**
4.3 Mission Implementation Stage (MI) **12**
4.4 Mission Operations Stage (MO) **14**
4.5 Post Mission Stage (PM) - Post-Operations and Preservation **19**

ANNEX A – QUALITY INDICATORS **21**

List of Tables

Table 1: Applicable Documents **2**
Table 2: Reference Documents **2**
Table 3: Assets to be preserved during the Mission Concept Stage **10**
Table 4: Assets to be preserved during the Mission Definition Stage **12**
Table 5: Assets to be preserved during the Mission Implementation Stage **14**
Table 6: Assets to be preserved during the Mission Operations Stage **18**
Table 7: Assets to be preserved during the Post Mission Stage **20**
Table 8: Quality Indicators **22**

1. INTRODUCTION

1.1 Intended Audience

The document is intended to assist data managers in Earth Observation data centres in applying the CEOS Best Practices for Long Term Preservation of Earth Observation Space data (www.ceos.org) to ensure EO mission data set assets preservation, curation, and valorisation for long-term accessibility and exploitation.

1.2 Background

In 2006, the European Space Agency (ESA) initiated a coordination action to share a common approach towards the long-term preservation of Earth Observation space data among all European and Canadian data holders and archive owners. A Long Term Data Preservation (LTDP) Working Group was formed in Europe in 2007 to define and promote a coordinated approach for long-term data preservation and curation of European Earth Observation space data assets. One of the outputs of the group consisted of the 'EO Preserved Data Set Content', a best practice document guiding Earth Observation data holders in their preservation activities [RD-1]. The 'CEOS Preserved Data Set Content' generated in the frame of the CEOS WGISS Data Stewardship Interest Group (DSIG), has evolved from the European document to become a global reference for Earth Observation data preservation.

1.3 Document Scope

This document identifies the EO mission data set assets content (i.e. data records and associated knowledge) that should be preserved to ensure long-term usability and exploitation of Earth Science data.

The document is intended to provide the content description (the “what”) for all the items of the EO mission data records and knowledge that should be preserved beyond the mission lifetime. It is intended as a guideline on how to use the content description list to support CEOS Best Practices associated documents.

The document is also intended to assist data managers in making sure that the recommended and mandatory content is collected, certified for completeness and quality at data set generation, during each mission stage, thereby providing the list of expected documents, content and quality information to be generated and preserved at each stage.

In accordance with the CEOS Best Practices, the composition of the PDSC varies by sensor category and needs to be tailored for the specific data set at hand, taking into consideration the designated community, preservation objective, requirements, quality information, metadata generation and dependencies.

The data manager shall tailor the PDSC to meet the needs of the specific mission, stating which data records and knowledge should be preserved during each phase of the Preservation Workflow in accordance with [AD-1] and maintain the Preserved Data Set Content inventory table with the data records, information, and software available under configuration, in accordance with [AD-2].

1.4 Applicable and Reference Documents

ID	Resource
[AD-1]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data - EO Data Stewardship Definition
[AD-2]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data – Preservation Workflow – March 2015 Version 1.0
[AD-3]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data –EO Data Preservation Guidelines, CEOS/WGISS/DSIG/EOPG, June 2015
[AD-4]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data – Generic Earth Observation Data Set Consolidation Process version 1.0 March 2015
[AD-5]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data – Persistent Identifier
[AD-6]	CEOS EO Data Purge Alert Procedure, http://wgiss.ceos.org/purgealert/
[AD-7]	Quality Assurance Framework for Earth Observation - Guidelines Framework (QA4EO). www.qa4eo.org.

Table 1: Applicable Documents

ID	Resource
[RD-1]	EO Preserved Data Set Content v 4.0, LTDP-GSEG-EOPG-RD-11-0003, July 2012
[RD-2]	ISO 14721 - OAIS standard (ISO reference model for Open Archival Information System) Pink Book, Consultative Committee for Space Data Systems, Greenbelt, MD. August 2009.
[RD-3]	European Cooperation for Space Standardization (ECSS), http://www.ecss.nl/
[RD-4]	Producer-Archive Interface Methodology Abstract Standard (PAIMAS), 05/2004, CCSDS 651.0-M-1
[RD-5]	Producer-Archive Interface Specification (PAIS), CCSDS 651.1-R-1, 02/2012

Table 2: Reference Documents

2. HOW TO USE THIS EO PDSC

The need for accessing historical Earth Observation data information has greatly increased, driven by long term scientific and environmental monitoring.

This document is meant to provide assistance to the practical implementation at working level of [AD-1] to [AD-7], providing recommended guidelines in reply to:

- the "what" dimension
- the "when" dimension
- the "quality" dimension
- the "preservation metadata" dimension
- the "how" dimension

2.1 "What" Dimension

This document has undergone a significant public review of the "what" dimension, i.e. the content specification of what is mandatory to preserve beyond the mission lifetime (i.e. the measurements for which the instrument was designed for), either raw data (as acquired by the satellite and recorded at the stations or received via Third Parties), or otherwise, global or higher level mission products when systematically generated and/or reprocessed as part of the mission requirements.

It identifies all additional information required to correctly understand and interpret the primary data, including in particular ancillary data (e.g. spacecraft ephemeris information, attitude, etc.), auxiliary data (required to process the telemetry payload data to generate the nominal mission products), CAL/VAL databases, whenever available (including processing/reference validation data sets) and mission-related documentation, including description of mission products and of the algorithms needed to obtain them.

The detailed list is provided in Chapter 3.

2.2 "When" Dimension

The experience with historical mission recovery has underlined the need to ensure that the mission asset content is qualified as fit for purpose for the long term preservation, during the mission lifetime, in accordance with quality certifying processes. The "when" dimension described in Chapter 3 is intended to assist data managers in ensuring that the recommended and mandatory content is collected, certified for completeness and quality at data set generation, during each mission stage, thereby providing the list of expected documents, content and quality information.

2.3 "Quality" Dimension

The need to preserve EO mission data assets indefinitely has led to the establishment of several CEOS Best Practices on Long Term Preservation of Earth Observation Space Data [AD-1] to [AD-6]. To guarantee that the preserved data set is "fit for purpose", it is mandatory to preserve its quality information. This is the objective of the Quality Assurance for Earth Observation (QA4EO) framework [AD-7] developed by the CEOS Working Group on Calibration and Validation (CEOS WGCV).

The Quality Assurance for Earth Observation (QA4EO) framework aims to provide EO data users with sufficient but simple information to enable them to evaluate the fitness for purpose of data/information for their applications, while also facilitating harmonisation and interoperability of

data sources. The key principle is stated in QA4EO Study results as: Data and derived products shall have associated with them an indicator of quality to enable users to assess their suitability for particular applications, i.e. their "fitness for purpose".

This can be expanded further, requiring that all EO data and derived products have associated a documented and fully traceable quality indicator, where:

- A Quality Indicator shall provide sufficient information to allow all users to readily evaluate the "fitness for purpose" of the data or derived product.
- A Quality Indicator shall be based on a documented and quantifiable assessment of evidence demonstrating the level of traceability to internationally agreed (where possible SI) reference standards.

A Quality Indicator may be a number, set of numbers, graph, uncertainty budget, or a simple "flag" (see Table 8 for a, non-exhaustive, list of Quality Indicators).

To address this, QA4EO contains a set of guiding principles, supported by a suite of "key guidelines" based on existing best practises [AD-7].

However, the concept of a quality indicator is of limited use for the purposes of deciding which one needs to be preserved, as a quality indicator appropriate for one set of users, may not apply to all users who might need different indicators.

Instead, in this document, the concept of quality information is defined. This represents the information needed to define a quality indicator, i.e. to assess the fitness for purpose of the EO data records. This information is part of the Preserved Data Set Content specification and this document indicates where the quality information should be found and at which stage of the mission lifetime (chapter 3).

2.4 "Preservation metadata" Dimension

Preservation metadata is defined as the metadata information that the data manager and steward need in support to the digital preservation process, stewardship and curation objectives as defined by the CEOS Best Practices in [AD-1] to [AD-6]. According to [RD-4] preservation metadata shall be generated during the life cycle of the asset to be preserved.

There are different types of descriptive metadata: domain specific, administrative (including rights and permissions), technical, documenting digital provenance, documenting relationships and links in the preservation repository.

2.5 "How" Dimension: Tailoring the PDSC

In accordance with the CEOS Best Practices, the composition of the PDSC varies by sensor category and needs to be tailored for the specific data set at hand, taking into consideration the designated community, preservation objective, requirements, quality information, metadata generation and dependencies.

The data manager shall tailor the PDSC to meet the needs of the specific mission, stating which data records and knowledge should be preserved during each phase of the Preservation Workflow in accordance with [AD-1] and maintain the Preserved Data Set Content inventory table with the data records, information, and software available under configuration, in accordance with [AD-2].

This tailoring should involve mission experts (e.g. instrument designers, quality working groups), but also the data end user communities, to ensure that their needs have been taken into account. The

tailored document should have a defined owner and should be kept under review throughout the mission, at a minimum at the end of each mission stage.

Procedures must be in place to ensure that all quality information identified in the PDSC as being needed is in fact saved in the correct place. A common issue for historical missions was that significant information was captured in less formal ways, making it almost impossible to retrieve afterwards.

These procedures should include acceptance of a document being conditional on the quality information being complete, review of the quality information at milestones, and the transfer of all quality information to archives at the appropriate time.

Should some quality information be required but not present (e.g. because it is recorded in a different document than that specified by the PDSC), then the tailored PDSC should be updated to reflect the actual situation.

All quality information must be stored using the processes described in the Preservation Workflow CEOS Best Practices [AD-2].

The following requirements should apply for the tailoring:

- **R01: The PDSC document should be tailored for each mission and instrument.**
- **R02: The PDSC tailoring should be reviewed at least at the start of each mission stage.**
- **R03: The tailored PDSC should be made available to the designated community for review and feedback.**
- **R04: All quality information identified in the tailored PDSC should subsequently be documented and saved, with clear procedures in place to ensure this.**
- **R05: To facilitate checking that the PDSC has been complied to, all items of quality information should be given an identifier specifying the row that they correspond to.**
- **R06: To ensure that quality information required is available, the PDSC should be used to define deliverables for (sub) contracts.**
- **R07: The project office should maintain a directory of the knowledge information, and specifically of the linkages between items.**
- **R08: A suitable tool shall be developed to record and allow traceability of linkages between items of quality information.**
- **R09: The tool shall be used to record the quality information and the linkages between items, for a given mission or instrument.**
- **R10: A copy of the knowledge information identified in the PDSC should be stored in the same archiving centre as the data records.**
- **R11: Whenever possible, an automatic tool shall be provided to allow tracing the knowledge information relevant for a given data record.**
- **R12: Effort shall be made to ensure that all documentation, tools, calibration data and other associated knowledge are free from any legal or commercial restriction.**

3. PRESERVED DATA SET CONTENT SPECIFICATION

The Preserved Data Set Content specification is intended to provide the content description (the "what") for all the items of the EO mission data records and knowledge that should be preserved beyond the mission lifetime.

In the document the term Mission is used generically and includes the concept of "Experiment", "Campaign", "Project", etc.

EO Missions/Sensors Dataset is defined as:

- **Data Records:** these include raw data and/or Level-0 data, higher-level products, browse images, auxiliary and ancillary data, calibration and validation data sets, and descriptive metadata;
- **Associated Knowledge:** this includes all the Tools used in the Data Records generation, quality control, visualization and value adding, and all the Information needed to make the Data Records understandable and usable by the Designated Community (e.g. mission architecture, products specifications, instruments characteristics, algorithms description, calibration and validation procedures, mission/instruments performances reports, quality related information). It includes all Data Records Representation Information, Packaging Information and Preservation Descriptive Information according to the OAIS information model (part of this information might be included in the descriptive metadata depending on the specific implementation).

3.1 Data Records

Data records are identified as:

1. Raw data[1]
2. Level 0 data (L0)
3. Level 1 (L1) to higher levels mission data products when generated as part of the mission requirements and/or reprocessed
4. Browses whenever generated
5. Ancillary data (spacecraft ephemeris information, attitude, etc.)
6. Auxiliary data (required to process the telemetry payload data to generate the nominal mission products)
7. Calibration and validation datasets[2] (needed to calibrate the satellite instruments and monitor data quality)
8. Metadata

[1] Raw data shall be preserved whenever conversion to Level 0 cannot be adequately certified.

[2] Including processing/reference validation data sets.

3.2 Tools

This includes:

1. L0 consolidation software[3]
2. Data processing software (for products generation from Level 0 to higher levels according to mission requirements)[4]
3. Quality control software
4. Data/products visualization tools
5. Value adding tools

3.3 Information

It is assumed that each document part of the "Information" generated and identified in one of the stages below is maintained and updated in the following stages according to mission evolution. Documents that might evolve are included below only in the first stage during which they are generated even if they are maintained and updated during the subsequent stages of the mission.

Mission or project related documentation is generally identified by:

1. Mission architecture documents describing purpose, scope and performances of the mission and of the on-board instruments, information regarding relevant orbits, platform position, attitude, ground coverage (acquisition footprint), head-roll-pitch.
2. Documents describing data and product format specifications.
3. Documents describing measurement requirements and/or measurement performances (theoretical models). Documents regarding instruments characteristics, performances and instrument description (physical implementations). Documents describing models and/or algorithms needed/used to obtain mission data and products, including specific/special cases, known errors and configuration necessities. In other words, all documents covering the conceptual environment, its implementation and its operations.
4. Reports concerned with measurement trends, failures, changes of performances, un-availabilities
5. Reports and outcomes from events such as: congresses, studies, communities and investigators concerned with models' review, algorithm changes, and Cal/Val changes affecting data processing chains.
6. Documents related to the process of data qualification: precision, numerical representations, formats, uncertainties, errors, adjustment/correction methods (e.g. Cal/Val procedures and documents).
7. Document related to workflows, work procedure, documentation three and bi-directional link
8. Scientific publications based on the data exploitation or relevant to them (properly linked to the data) and outreach material.
9. Administrative (Memorandum, Intellectual Property Rights, etc.)
10. Mission Data Records and Documentation Tree

---

[3] Whenever raw data are preserved.

[4] Data Processing Software could be maintained in operation to generate mission products or all products could be generated through a Bulk Processing Campaign and Software code and algorithms archived.

Mission documentation shall include Representation Information, Packaging Information and Preservation Descriptive Information according to OAIS Information Model [RD-4].

4. PRESERVED DATA SET CONTENT FOR EARTH OBSERVATION MISSIONS

The PDSC should be tailored appropriately for each mission/instrument. The tailoring of the PDSC should involve mission experts (e.g. instrument designers, quality working groups) and the designated user communities. The tailored document should have a defined owner, and should be kept under review throughout the mission lifecycle, at the end of each mission phase/stage. Procedures must be in place to ensure that all quality information identified in the PDSC as traceable and preserved. These procedures should include acceptance of a document being conditional on the quality information being complete, review of the quality information at milestones, and the transfer of all quality information to archives at the appropriate time

An Earth Observation space mission is generally divided into the following stages:

1. **Mission Concept (MC).** Defines the mission to a sufficient level to show the scientific value and technical feasibility. During this stage, identification of the science requirements by the Science study team and study scientist are carried out. Additional activities include the identification of a reference platform to be used in the preliminary system level studies. Feasibility verification documents, mission technology and programmatic estimates for the future mission stages are also generated. According to ECSS standards [RD-3] the Mission Concept stage can be identified as Phase A of mission design.

2. **Mission Definition (MD).** This stage is concerned with the mission scientific requirements detailed definition and the selection of technical solutions for system concept. During this stage, types of scientific instrument measurements (e.g. spectral analysis, temperature measurement, etc.) are identified and defined eventually combining existing sensors/instruments in different modes or with different scientific models. According to ECSS standards the Mission Definition stage can be identified as Phase B of mission design.

3. **Mission Implementation (MI).** According to Mission Definition results, this stage produces the detailed definition and implementation of the mission system and components: sensors/instruments; algorithms and their relationship in the frame of scientific domains; methods of measurement and any other context necessary to perform measures. Production, development testing and pre-qualification of selected critical elements and components lead to the conclusion of the technology development activities. According to ECSS standards the Mission Implementation stage can be identified as Phases C/D of mission design and implementation.

4. **Mission Operations (MO).** This stage identifies the operational timeframe of the mission being the period during which data are captured, algorithms are revised and improved, activities concerned with input analysis, calibration and validation of sensor/instrument as well as activities concerned with qualification of processed data are performed. According to ECSS standards the Mission Operations stage can be identified as the Phase E Operations till the end of mission lifecycle.

5. **Post Mission (PM).** This represents the ***Post-Operations and Preservation stages***. The Post Mission stage is usually identified according to current ECSS standards as the Phase F of a mission. In this document the Post Mission stage has been extended and augmented and mainly focuses on the archived data to accommodate the need to preserve them in the long term for further reuse and exploitation. The post mission stage starts with the satellite end of life (e.g. for an Earth Observation mission with the event of satellite disposal or failure). The Post Mission stage focuses on datasets (data and information) consolidation and appraisal, datasets reprocessing to align to the latest version, ground segment and media disposal (depending on specific mission), and data and associated information migration to a

long-term preservation environment. During the Post Mission stage, a limited set of functions (e.g. data discovery and access) is provided by the mission ground segment (still in operation) according to the adopted strategy and depending on mission requirements until its disposal and data migration to long-term preservation. This stage also focuses on historical data reuse and exploitation, on data and concerned information preservation against aging and technological changes, and on data curation and enrichment).

4.1 Mission Concept Stage (MC)

Rationale – Information produced during this stage provides a snapshot of the scientific and technical framework in which the mission was born. Mission and sensors requirements, assessment studies, technology readiness review and cost analysis are performed during this stage. Preserving this information – both for approved and not approved missions – would allow future users to have reference material for new missions' evaluation and definition. Traceability of this information is also useful to compare initial expectations to what was actually achieved by the mission and to understand which changes occurred between the pre-mission and the next stages.

ID	Type	Identification	Description	Quality Information	Notes
MC 1.1	Doc	Scientific Scenario and User Communities	Defines scientific scenario and expected goals. Also list Principal Investigator, designated user communities and third party actors.	Required uncertainty for services and applications, lifetime, data availability, data accuracy, data latency, revisit time, geographical coverage, spatial resolution.	
MC 1.2	Doc	Mission Requirement Document	Defines scientific mission and sensor requirements, processing methods, qualification, methods	Calibration plan and quality assessment plan for the mission. Uncertainty requirements for instrument product (e.g. radiometric/geometric uncertainty, coverage, revisit time, etc.) Justification for the design decisions (e.g. band selection)	Most information should be contained in the mission documentation, e.g. the Mission Requirement Document (MRD), Mission Operations Concept Document (MOCD) and Mission Description Document (MDD) according to ECSS or equivalent standards.
MC 1.3	Doc	Mission Operation Plan	Defines the plan on how the mission will be conducted	Plan for handling quality information	

Table 3: Assets to be preserved during the Mission Concept Stage

4.2 Mission Definition Stage (MD)

Rationale: The Mission Definition stage produces the entire mission and data detailed definition documents. It includes Sensor/Instrument requirements, characteristics, calibration methods, etc. Preserving this information is fundamental to understand changes that may have occurred over time while in operation.

ID	Type	Identification	Description	Quality Information	Note
MD 1.1	Doc	Mission Requirements Specifications	Defines mission requirements, mission space to ground functional and resource allocation and operational scenario. Contains the specifications for the verification and validation method for space to ground resources	Description of the quality information at a global (e.g. revisit times and mission products uncertainty) and at a subsystem level: instrument e.g. straylight, channel crosstalk, spatial sampling. FoV, observation mode, spectral channels.	Most information should be contained in the System Requirement Document and Justification File, System Functional Specification, and Design Definition File (DDF), Design Justification File (DJF) documents according to ECSS standards and equivalent
MD 1.2	Doc	Space to Ground segment ICDs	Defines the main systems / segments ICDs, system budget estimation and data flow.	Error Control (e.g. CRC) data latency, data rate, quality flags, packet lost/damaged, timeliness etc. for different scenarios (e.g. Near-Real-Time NRT, calibration mode, ground stations availability and relative position)	Most information should be contained in the space-to-ground interface control document (SGICD) according to ECSS standards.
MD 1.3A	Doc	Sensor / Instrument requirements	Defines the Sensor / Instruments requirements for design (e.g. bands, modes, performances, etc.).	Sensor uncertainty budget based on previous knowledge. Specification of uncertainty associated with optical properties e.g. noise, linearity, calibration accuracy, signal synchronisation, electrostatic protection, temperature and pressure range.	
MD 1.3B	Doc / Data Record	Sensor / Instrument characteristic	Characteristic for processing of acquired data, data processing model	Assessment of performance/ acceptability including uncertainty, linearity, sun-glint, straylight: Documented model descriptions, validation of model and software, version control. Validation by comparison with other models or reference data sets including simulated products and ground measurements.	This includes validation campaigns with in-situ products

MD 1.4A	Doc / Data Record	Sensor / Instrument qualification process	Qualification process for sensor, captured data, processed data.	Documented procedure for validation.	
MD 1.4B	Doc / Data Record	On-ground calibration and characterisation plan	Calibration requirements	Identification of reference standards, pre-flight calibration methods, re-calibration intervals Uncertainty aims	Pre-launch calibration campaign includes: • Optical Tests • Thermal Test • External Calibration Test • Field Of View determination
MD 1.4C	Doc / Data Record	Ground/Ocean calibration reference and scientific base	Calibration requirements	Traceability to International System of Units (SI) via international reference standards: Procedures, calibration certificates, traceability statement, and uncertainty analysis.	Should include description of these sites, accuracy, stability of the site conditions.
MD 1.5	Doc	Processing algorithms and data format specification	Defines: Mathematical models and algorithms for mission data processing; Auxiliary and ancillary data orchestration; Data and Products format requirements and standards.	Documented descriptions of mathematical models and algorithms for mission data processing; including: Assessment of performance / acceptability. Peer reviewed papers Simulation for validation results. Validation by comparison with test datasets Validation of performance simulator. Auxiliary and ancillary data orchestration. Data and Products format requirements and standards including: Metadata specifications (including quality information/parameters) Naming conventions Version controls specification	Should define what validation evidence is required to accept any product.

Table 4: Assets to be preserved during the Mission Definition Stage

4.3 Mission Implementation Stage (MI)

Rationale: Preserving all the information produced during the Mission Implementation stage is needed to understand procedural impacts relative to instrument, algorithm and product implementation. Data acquired during the calibration and validation campaigns of the instrument under construction (e.g. in a laboratory or dedicated campaigns) is of critical importance as a reference for the future use of the data.

ID	Type	Identification	Description	Quality Information	Note
MI 1.1	Doc	Mission Design (Space and Ground Segment)	Defines mission requirements specification and implementation design.	Clear identification of technical procedure. Record of decision made during implementation	Most information should be contained in the System and Subsystems Requirement Documents and Justification Files, System Technical and Functional Specifications including Interface Requirements, Design Definition Files and Design Justification Files according to ECSS standards.
MI 1.2A	Doc	Detailed Space to Ground Segment Operations Concept and implementation	Defines the detailed Space to Ground operational implementation and any contingency procedure/plan needed	Recording procedure for assuring the data quality. Storing of diagnostic information received.	Most information should be contained in the consolidated Mission Operations Concept and Space to Ground Technical Budget documents according to ECSS standards.
MI 1.2B	Doc	Data Handling	Data Capture and handling	Clear identification of technical procedure	
MI 1.2C	Doc	On Board Processing	On board processing	Algorithm description and software validation	
MI 1.3	Doc	Sensor/Instrument Design and Implementation	Defines the Sensor/Instrument platform design and implementation and its performances Platform and instrument design implementation/test, budges performances.	Testing results including uncertainty. Uncertainty budget with supporting evidence (from on ground characterisation). Uncertainty combination, covariance.	In this item it is possible to include the information of the other relevant subsystem with a direct impact on the mission data performances (e.g. attitude and orbit determination subsystem).
MI 1.4	Doc / Data Records	Validation and Calibration	Independent validation and calibration campaign method, data validation activities with simulated data.	Calibration results, uncertainty budget with supporting evidence, traceability to SI validation results	This includes the pre-flight calibration/validation campaign and should focus on calibration rather than validation.

MI 1.5A	Doc	Ground Processors Design, Algorithms Implementation and Supporting Information for data processing.	Defines the design and implementation of the ground data processors and the algorithm. It includes also supporting information for data processing (e.g. ancillary, auxiliary data description & orchestration, etc.)	Algorithm description and software validation for all software used on ground and on board. Metadata and naming conventions, version control	
MI 1.5B	Notes/ Papers	Technical Notes Scientific Paper Peer Review	Version Control	Algorithm description and software validation for all software used on ground and on board. Metadata and naming conventions, version control	
MI 1.6A	Doc	Products Specifications	Provides a detailed description of product and their characteristic	Description of uncertainty/quality indicators and method to provide uncertainty to different users.	
MI 1.6B	Doc	Data Format Specifications	Contains information that will allow the user to read and use the data.	Data format naming conventions, performances of compression algorithm, quality indicator specification	
MI 1.6C	Doc / Data Records	Supporting Information for processing	Ancillary and auxiliary definition and identifications, orchestrations	Appropriate quality indicator for ancillary/auxiliary data to be used in the mission operations stage with the relevant metadata	
MI 1.7	Doc / Data Records	Qualification Process	Detailed qualification methods and data	Assessment of performances/acceptability	

Table 5: Assets to be preserved during the Mission Implementation Stage

4.4 Mission Operations Stage (MO)

Rationale: Data acquired during the Mission Operation stage is the concrete heritage that the mission will leave to future generations. The Mission Operation stage provides the effective data that will be analysed by the scientific community and that will the core of the mission preservation objective. The data also serve public administration and commercial applications, which depend on reliable, sustainable data availability to fulfil their public task and to set up viable business cases. Software related to this mission stage needs to be preserved in order to use, process and exploit data in the future. Documents also need to be preserved to have a comprehension of the data itself and to perform mission results qualification.

ID	Type	Identification	Description	Quality Information	Notes
MO 1.1	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 services requirements & performances during the operations stage.	Clear identification of technical procedure	Phase E1 and Phase E2 ECSS standards equivalent
MO 1.2	Doc	Sensor Ground Segment Operations Plan	Defines the actual implementation of the end-to-end mission operations.	Uncertainty budget with supporting evidence	
MO 1.3	Doc / Data Records	Mission Operations Acquisition Plans and Reports	Describes the mission sensor acquisition plans and reports.	Availability of data, data quality, model evolution, calibration parameters evolutions, geo- location performance, data anomalies	Phase E1 and Phase E2 ECSS standards equivalent
MO 1.4	Data Records	Raw/Level 0	Raw or Level 0 data from the sensor or instrument data packets	Completeness of data, timeline, Certification of L0 processing (unless stored as raw). Noise – SNR & SD of the data	Raw data shall be preserved whenever conversion to Level 0 cannot be adequately certified
MO 1.5A	Data Records	Level 1	Processed image data L1 products	Associated uncertainties and evidence. Processing algorithm recorded and validated. Reference to calibrations, traceability. For geometrically located area, geometric alignment and resampling	
MO 1.5B	Data Records	Level 2	Processed image data L2 products and higher	Associated uncertainties and evidence. Processing algorithm recorded and validated. Reference to calibrations, atmospheric corrections, traceability. Reference	

				to validation where relevant	
MO 1.6	Data Records	Browses/Images	Browse Digital Catalogue	No specific quality information is needed for the browse images.	Whenever generated
MO 1.7	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 instrument parameters (e.g. optical response	Quality flags and performance parameters e.g. orbit accuracy, temperature stability	
MO 1.8	Data Records	Auxiliary Data	Band/Multispectral/ Band-by-band parameters for algorithms, Non linearity correction factors, Error/Failure/Gap correction factors, Calibration curve/Factors, 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/geode model DEM, etc.	Associated uncertainties and evidence where appropriate otherwise performances flags and parameters e.g. drift. Sensitivity coefficients for L1 and L2 data to their parameters. Date range for auxiliary file version.	Required to process the telemetry payload data to generate the nominal mission products
MO 1.9	Doc/ Data Records	Calibration and validation data	Cal/Val data acquired during mission operations (optical/radiometric stability, Instrument availability, Internal calibration, Optic pointing pattern, etc.)	In-flight calibration reports, uncertainly with evidence, version report, instrument anomalies Parameters evolution (degradation model, DS, pixel response linearity…) Instrument validation: SNR validation, absolute and relative radiometric vicarious calibration, MTF, geolocation, L2 products… Validation reports, satellite uncertainties.	Cal/Val data acquired during mission operations through validation campaigns run to calibrate the satellite instruments and monitor data quality. It includes processing/reference validation data sets. Includes also related documentation (e.g. reports).

MO 1.10	Doc/Data Records	Quality Parameters	PA/QA of instrument, raw data and products	Assessment of performance/accept ability	
MO 1.11	Doc/Data Records	Metadata	Metadata Digital Inventory	No specific quality information is needed for the metadata	The metadata can be generated from auxiliary, ancillary and similar data and can always be recovered if appropriate procedures are set in place.
MO 1.12	SW Code	Level 0 consolidation		Algorithms and software verification / Validation, version control	
MO 1.13	SW Code	Data Processing Software	Instrument processing algorithms, context and source codes, testing context	Algorithm description. Algorithms and software verification / Validation, version control	
MO 1.14	SW Code	Quality Control Software		Algorithms and software validation, Algorithms and software verification / Validation, Version control	
MO 1.15	SW Code	Visualization Tools	Processing and visualizing tools	Software validation and version control Algorithms and software verification / Validation, Version control	
MO 1.16	SW Code	Value-Added Software		Software validation and version control Algorithms and software verification / Validation, Version control	
MO 1.17	SW Code	Data/ image processing	Packed telemetry, PUS, CCSDS, Instrument source packet, product formats, and storage formats.	Software validation and version control, software developments	
MO 1.18	Doc	Product qualification and quality assurance	Defines the product qualification process	Assessment of performance/accept ability based on relevant	

		monitoring reports	outputs.	quality parameters such as uncertainty levels, flags etc.	
MO 1.19	Doc	Sensor/Instrument evolution and history records	Describes any instrument event that might affect data quality (e.g. upgrading, downgrading, LUTs). It includes also known-errors and limits of sensors/instruments.	Instrument timeline Documented supporting evidence for decisions	
MO 1.20	Doc	Referred publications and papers	Referred publications, articles and technical notes clearly referencing the used datasets.	No specific quality information has been requested	Any future publication should be enforced to provide clear reference to the utilized dataset.
MO 1.21	Doc	Tandem and/or combined campaigns, comparisons	Data and reports	Uncertainty budgets with supporting evidence Comparisons report following QAE4EO Guideline 7	
MO 1.22A	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	Evidence of participation in appropriate comparisons. Comparison report following QA4EO Guidelines 4 and 7	
MO 1.22A	Doc / Data Records	Cross- campaign, cross-comparisons and cross- calibration activities documentation n and Data	Describes the cross campaign scenario and operational context. Also describes any cross-calibration activities	Evidence of participation in appropriate comparisons. Comparison report following QA4EO Guidelines 4 and 7	
MO 1.23	Doc	Data Access Policy	Describes the data access policy for mission in the operational stage.		

Table 6: Assets to be preserved during the Mission Operations Stage

4.5 Post Mission Stage (PM) - Post-Operations and Preservation

Rationale: After the end of a mission, datasets acquired during the operational stage need to be consolidated and aligned to the latest available version of the processors and/or improved version. All the evolution activities carried out in the previous stages and the changes to the data and associated information are properly assessed and consolidated during this stage for end-to-end consistency/coherency/provenance based on the documentation produced and preserved in the previous stages. During this stage the user communities will still need to analyse and process data. Enhanced algorithms and processors improvements could be implemented to improve data exploitation and processing performances.

ID	Type	Identification	Description	Quality Information	Note
PM 1.1A	Doc	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Processing	Processing and/or Calibration change including provenance e and context	Algorithms and software validation, version control Clear description of motivation for reprocessing and improvements gained	Level 0 data consolidation should be certified in this stage and in such a case raw data could be disposed.
PM 1.1B	Doc/Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Ancillary, Auxiliary	Updated Ancillary, Auxiliary	Associated uncertainties and evidence, version control	
PM 1.1 C	Doc/Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. PA/QA	Quality information updated as part of reprocessing	Assessment of performance/ Acceptability	
PM 1.2	Data Records (Reproce ssed data set)	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. L0, L1, L2	Reprocessed data & products	Associated uncertainties and evidence, version control	
PM 1.3	Doc	Data consolidation & reprocessing strategy, implementation plans and consolidated/ reprocessed data. Data/Image processing	Instrument processing algorithms,	Algorithm and software validation, version control	
PM 1.4	Data	Data consolidation & reprocessing strategy, implementation plans and consolidated/ reprocessed data. Browse Metadata	Metadata Inventory		

PM 1.5	Docume ntation	Referred publications and papers	Referred publications, articles and technical notes clearly referencing the used datasets.	PID	
PM 1.6	Doc	Historical Data Access Policy	Describes the data access policy for the historical mission in the Preservation stage.		
PM 1.7	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 includes also the appraisal of the mission datasets (i.e. their value).	Summary of quality information approach within mission / instrument	Generated starting from information collected in the previous stages.

Table 7: Assets to be preserved during the Post Mission Stage

ANNEX A – QUALITY INDICATORS

Term	Definition
Quality Indicator	A means of providing a user of data or derived products (resulting from processing of data) with sufficient information to assess its suitability for a particular application. This information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally SI), but can be presented as numeric or a text descriptor, providing the quantitative linkage is defined. For many missions this will mean a documented and complete uncertainty budget (see QA4EO-QAEO-GEN-DQK-006), with quantitative evidence of traceability (see QA4EO-QAEO-GEN-DQK-007), though for some applications it will be sufficient to describe biases to agreed references or other sensors. The QI is likely to be presented as a report.
Uncertainty	Non-negative parameter characterising the dispersion of the quantity values that are being attributed to a measure and (quantity), based on the information used. A measure of the standard deviation of the probability distribution for the measure. Where possible this should be derived from an experimental evaluation but can also be an estimate based on other information, e.g. experience. Uncertainty evaluation should start by identification of a measurement equation. The sensitivity of the determined measure and to each effect in the measurement equation can be calculated either through partial derivation of the measurement equation, or through experimental investigation of the effect. The different uncertainty contributions are listed in an "uncertainty budget" and combined in quadrature. The standard uncertainty can then be, as appropriate, expanded with a coverage factor, for example to obtain a 95 % confidence level. The analysis of uncertainty is described in QA4EO-QAEO-GEN-DQK-006.
Traceability	Property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. In practice traceability is obtained by a series of comparison each of them involves calibration standard at one level in the chain using a standard at a higher level. Ideally traceability will lead back to the SI, through a National Measurement Institute. For example an irradiance-mode radiometer may be calibrated against a standard irradiance source (lamp), which was calibrated against a primary irradiance source at a National Measurement Institute (a blackbody), whose irradiance properties were known due to a filter radiometer (effectively an absolute pyrometer), which was calibrated against the primary radiometric reference (the cryogenic radiometer) and thus to SI. At each stage in the traceability chain there needs to be documented evidence of traceability, in the form of a calibration certificate, along with documented procedures and validation. This is described in QA4EO-QAEO-GEN-DQK-007.
Sensitivity (Coefficients)	This determines how sensitive the measure and (e.g. a L1 or L2 data product) is to any particular source of uncertainty. Some sensitivity coefficients can be calculated by differentiating the measurement equation (e.g. an inverse square law behaviour makes the sensitivity of irradiance to distance a factor of two: a 1 % change in distance, makes a 2 % change in irradiance). Other sensitivity coefficients are determined experimentally, e.g. by changing the temperature of the sensor, it is possible to determine how sensitive the signal on that sensor is to temperature changes. It may also be necessary to determine the sensitivity of model results to changes in the assumptions of that model.

Term	Definition
Calibration	Assessment of the correct values to the instrument's measurement scale by comparison with a reference standard of higher accuracy (higher level at the traceability chain). For example an instrument's spectral radiance responsivity is calibrated by putting it in front of a reference radiance source, whose radiance is determined traceable. Every step of a calibration chain needs documentation, including reference standard properties and suitability (see QA4EO-QAEO-GEN-DQK-003), documented procedures (see QA4EO-QAEO-GEN-DQK-002) and evidence of traceability (QA4EO-QAEO-GEN-DQK-007).
Reference Standard	Realisation of the definition of a given quantity, ideally with a stated uncertainty, which can be used as a reference; it can be individual or community defined. A reference standard can be an artefact such as a lamp or a reflectance tile of known and certified irradiance or reflectance and associated uncertainty. The measurements against a reference standard are calibrations. A reference standard can be a calibrated instrument that is compared with the test instrument. A reference standard might refer to the calibration sites that had been previously characterised and are monitored from the ground. In all cases a reference standard needs to have known properties, with formal calibration, and must be used within its range of validity and in an appropriate manner. This process must be documented (see QA4EO-QAEO-GEN-DQK-003).
Validation	Confirmation that the performance (of an instrument, algorithm, or software) that fits the intended purpose. Performance of instruments and software can be validated by testing performance against known standards, and formal auditing processes. See QA4EO-QAEO-GEN-DQK-005 (for software and algorithms).
Supporting evidence	Documentation describing how a process was carried out and its traceability. Includes calibration certificates, documentary procedures, records of software validation, records of traceability. See QA4EO-QAEO-GEN-DQK-007 for a list of suitable evidence of traceability.
Comparisons	Organised peer-to-peer comparisons, where different sensors/calibration laboratories/etc. measure the same reference standard and results are compared with each other in a formal way. A formal comparison will follow the procedure described in: QA4EO-QAEO-GEN-DQK-004.

Table 8: Quality Indicators

Earth Science Data and Information System Project, Code 423

423-SPEC-001

NASA Earth Science Data Preservation Content Specification

Original, Change 01

January 2013

Expires: January 2018

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland

This page intentionally left blank.

NASA Earth Science Data Preservation Content Specification

Prepared By:

Signature obtained on file	**December 12, 2012**
H. K. Ramapriyan ESDIS Assistant Project Manager GSFC Code 423	Date
Signature obtained on file	**December 28, 2012**
John F. Moses ESDIS DAAC Engineer GSFC Code 423	Date

Approved By:

Signature obtained on file	**December 12, 2012**
Dawn R. Lowe ESDIS Project Manager GSFC Code 423	Date
Signature obtained on file	**January 3, 2013**
Martha E. Maiden Program Executive for Earth Science Data Systems NASA Earth Sciences Division NASA HQ-DK000	Date

Goddard Space Flight Center
Greenbelt, Maryland

This page intentionally left blank.

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, and changes to this document shall be made by document change notice (DCN) or by complete revision.

Any questions should be addressed to:

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

This page intentionally left blank.

Abstract

This document defines the contents of data, metadata and associated documentation to be preserved beyond the life of missions funded 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 not on the implementation or representation (i.e., "how") of the content items. The content items are divided into eight categories: Preflight/Pre-Operations Calibration, Science Data Products, Science Data Product Documentation, Mission Data Calibration, Science Data Product Software, Science Data Product Algorithm Input, Science Data Product Validation and Science Data Software Tools. Items are described under each of these categories along with rationale for requiring their preservation.

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

This page intentionally left blank.

Change Record Page

ISSUE	EFFECTIVE DATE	PAGES AFFECTED	DESCRIPTION	CCR APRROVED DATE
Original	11/23/2011	All	CCR 423-ESDIS-4	11/21/2011
Original, CH 01	01/03/2013	Title, vii, ix, 1-1, 1-2, 2-1, 3-1, 3-2, 3-5, 3-7	CCR 423-ESDIS-28	11/26/2012

This page intentionally left blank.

List of Affected Pages

Page No.	Revision	Page No.	Revision	Page No.	Revision	Page No.	Revision
Title	Original, CH01	3-7	Original, CH01				
i	Original	3-8	Original				
ii	Original	AB-1	Original				
iii	Original	AB-2	Original				
iv	Original						
v	Original						
vi	Original						
vii	Original, CH01						
viii	Original						
ix	Original, CH01						
x	Original						
xi	Original						
xii	Original						
1-1	Original, CH01						
1-2	Original, CH01						
1-3	Original						
1-4	Original						
2-1	Original, CH01						
2-2	Original						
3-1	Original, CH01						
3-2	Original, CH01						
3-3	Original						
3-4	Original						
3-5	Original, CH01						
3-6	Original						

This page intentionally left blank.

Contents

PREFACE III

ABSTRACT V

CHANGE RECORD PAGE VII

LIST OF AFFECTED PAGES IX

CONTENTS X

1. INTRODUCTION 1

1.1 Scope and Background 1

1.2 Organization 3

2. REFERENCES 1

2.1 Related Documents 1

2.2 Other References 1

3. CONTENT SPECIFICATIONS 1

3.1 Preflight/Pre-Operations Calibration 1

3.1.1 Instrument Description 1

3.1.2 Preflight/Pre-operational Calibration Data 1

3.2 Science Data Products 1

3.2.1 Raw Data and Derived Products 1

3.2.2 Metadata 2

3.3 Science Data Product Documentation 2

3.3.1 Product Team 2

3.3.2 Product Requirements and Designs 2

3.3.4 Product Generation Algorithms 3

3.3.5 Product Quality 4

3.3.6 Product Application 4

3.4 Mission Data Calibration 4

3.4.1 Calibration Method 4

3.4.2 Calibration Data 5

3.5 Science Data Product Software 5

3.6 Science Data Product Algorithm Inputs 6

3.7 Science Data Product Validation 6

3.8 Science Data Software Tools 7

3.9 Deliverables Checklist 7

ABBREVIATIONS AND ACRONYMS 1

1. Introduction

1.1 Scope and Background

One of NASA's strategic objectives is to "Study Earth from space to advance scientific understanding and meet societal needs". 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. The data resulting from NASA's missions 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 missions' 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 missions and observations, it is essential to preserve the previous mission 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 missions. While NASA is not legislatively mandated to preserve data permanently as other agencies are (e.g., USGS, NOAA and NARA), it is essential for NASA to preserve all the data and associated content beyond the lives of NASA's missions to meet NASA's near-term objective of providing access to data and services for active scientific research. Also, NASA has to ensure that the data and associated content are preserved for transition to permanent archival agencies. To fulfill this responsibility, identification of the specific content items that need to be preserved from each of NASA's missions 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 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 Open Archival Information System (RM-OAIS). Subsequent activities continue to expand through a range of related interests that reach toward more practical guidance for developing agency standards. They include provider-archive interchange recommendations (2004) and 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. **CH01**

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 missions 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 preserved as part of the provenance and context of the mission datasets.

The Reference Model for 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 communities should become the Designated Community 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 will help in developing standard representative information requirements for Earth Science data.

At this time, there are no standards that address content to be preserved for the benefit of future Earth science investigations, especially for future long time series climatological studies. 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 multiple organizations and disciplinary areas. Coordination among organizations will be needed to arrive at common standards. Adoption of this document as NASA's specification of content for preserving Earth science data will be a first step in such coordination. It is possible that this specification will need to be updated as a broader standard is developed, but this specification should serve the interim purpose of gathering preservation content from past and current missions and of ensuring that future mission planning includes preservation of the content items needed in the long-term.

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

approaching the end of their lives; and

- Recent work by ESDIS Project staff as participants in the Earth Science Information Partner (ESIP) Federation's Data Preservation and Stewardship Cluster (ESIP, 2011)

The USGCRP workshop had participants representing a wide range of scientific disciplines. The participants developed a number of 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, a number of content items were identified as important for preservation.

The ESDIS Project staff has worked with the EP-TOMS, ICESat GLAS and Aura HIRDLS 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.

The ESDIS Project staff has been active in the ESIP Federation's Data Preservation and Stewardship Cluster in proposing and developing an emerging standard for Provenance and Context Content. Representatives from several U.S. agencies are involved in this cluster, including NOAA and USGS. While developing such a standard and having it approved through an international standards body is a prolonged process (generally a few years), the content items identified so far are sufficient to provide a preliminary version of a specification for use within NASA.

1.2 Organization

The content specification is organized into the following eight categories: Preflight/Pre-Operations Calibration, Science Data Products, Science Data Product Documentation, Mission Data Calibration, Science Data Product Software, Science Data Product Algorithm Input, Science Data Product Validation and Science Data Software 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. In addition, to allow for differences among flight projects/instrument teams in approaches to packaging these contents into their deliverables, it is necessary to develop and deliver a "meta-document" which provides details of which deliverables cover which content items. This is discussed in Section 3.9. The description of each specification requirement is expected to evolve over time. The ESDIS Project configuration management process will be used to manage changes.

The focus of this document is intentionally on the content (i.e., the "what") of items that must be preserved rather than on the method or representation (i.e., the "how"). Some variation is expected in the degree with which these requirements are met depending on the phase of the

mission at which the requirements are addressed. The eight specification categories are intended to be comprehensive such that each is required to be addressed for a particular mission/instrument. However, the individual sub-items within the item descriptions are examples, and as such do not necessarily apply to every mission/instrument since NASA missions cover a broad range of sensor types, data and objectives. When applied to older missions where the projects have ended and the principals are not accessible, these requirements may not be satisfied fully while missions that are still in planning should be able to meet the requirements fully. The details of implementation shall be worked out while developing Interface Control Documents between the mission Project and the NASA ESD-assigned Data Center, which shall be held under configuration control by the ESDIS Project.

Not all of the data or information needed to satisfy a particular specification requirement may be physically archived at the assigned NASA Data Center. Some ancillary algorithm inputs (e.g., model data) may be archived by a partner agency such as USGS or NOAA. NASA sponsored documentation is archived by the NASA Scientific and Technical Information Program. The NASA STI program has provisions for handling material that is identified as having ITAR restrictions, or proprietary, is copyrighted and/or patent protected. If there are unpublished release notes, ancillary data and/or software that are ITAR restricted, proprietary, copyrighted or patent protected then one of the ESD Data Centers will need to provide the required safeguards for archive and distribution services. When providing these items to the Data Centers the mission Projects should include any distribution restrictions and expiration dates for those restrictions.

2. References

2.1 Related Documents

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 **CH01**
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
423-RQMT-003	NASA, 2011: Metadata Requirements – Base Reference for NASA Earth Science Data Products, (November 2011). **CH01**
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]

2.2 Other References

ESIP, 2011: Preservation and Stewardship Cluster of the Earth Science Information Partners Federation, http://wiki.esipfed.org/index.php/Preservation_and_Stewardship.

This page intentionally left blank.

3. Content Specifications

3.1 Preflight/Pre-Operations Calibration

3.1.1 Instrument Description

- Item Description: Documentation of Instrument/sensor characteristics including pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration (geo-location offsets), noise characteristics, etc.). For example, components of documentation 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 mission/Project documentation plans, this information may be found in a range of types of technical reports or (as in the opposite extreme), one over-arching document such as a user guide.
- Rationale: Needed for users to understand how the instrument operates. Documentation of measurements made before deploying instruments in space (or *in situ*) will help establish a baseline and help users understand changes that may have occurred over time while in operation.

3.1.2 Preflight/Pre-operational Calibration Data

- Item Description: 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), noise characteristics, etc.).
- Rationale: Measurements made before deploying instruments in space (or *in situ*) will help establish a baseline and help users understand changes that may have occurred over time while in operation.

3.2 Science Data Products

3.2.1 Raw Data and Derived Products

- Item Description: Raw data are data values at full resolution as directly measured by a spaceborne, airborne or *in situ* 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 at http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-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. **CH01**
- Rationale: Preservation of raw data (e.g., Level 0 data or Level 1a products) is required

for regeneration of any higher-level products in case errors are discovered or better atmospheric transmission/absorption/reflectance or scattering models become available in the future. It is important to preserve either the means of regenerating the higher-level products or the products themselves to ensure reproducibility and verifiability of scientific results.

3.2.2 Metadata

- Item Description: Information about data to facilitate discovery, search, access, understanding and usage associated with each of the data products. Metadata shall follow standards described in "Metadata Requirements – Base Reference for NASA Earth Science Data Products" (NASA, 2011).
- Rationale: Data cannot be located or obtained without discovery, search and access metadata. Data cannot be used without metadata needed for understanding and usage. For example, in some production environments it becomes critical that the granule-level metadata indicates which version of software was used to produce a particular granule. It is also important to include provenance information in granule level metadata to enable users to determine all inputs and parameters used in generating the granule.

3.3 Science Data Product Documentation

3.3.1 Product Team

- Item Description: Names of key science team leads and product team members (development, help desk and operations), roles, performing organization, 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.
- Rationale: It is important to know who were responsible for the 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 product quality/validity and consistency, and their publications may provide relevant insights. It is important to capture from the product team any comments about the products before the individuals move on to other activities and become unavailable to provide help to users.

3.3.2 Product Requirements and Designs

- Item Description: Requirements and designs for each science data product, either explicitly or by reference to the requirements/design documents. Product requirements and designs should include content, format, latency, accuracy and quality.
- Rationale: Explains expectation about the products when the project was initiated. Useful to compare with what was actually accomplished (as recorded in validation documents). Documentation of the product format is essential for a user to be able to use a product.

CH01

3.3.3 Processing and Algorithm Version History

- Item Description: For all products held in the archive, documentation of processing history and production version history, indicating which versions were used when, why different versions came about, and what the improvements were 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. Granule level metadata should indicate which version of software was used for producing a given granule. 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 granules 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 product granules that were used in scientific publications is essential to the scientific method.

3.3.4 Product Generation Algorithms

- Item Description: Detailed discussion of processing algorithms, outputs, error budgets and limitations 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 product (e.g., contained in peer-reviewed papers, in some cases supplemented by thematic information introducing the data set or derived product) - geo-location, radiometric calibration, geophysical parameters, sampling or mapping algorithms used in creation of the product, algorithm software documentation, & 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).
    - Description of the output data products at a level of detail to determine if the product 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: In order for any 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 product delivered to the Data Centers so that users of a particular version know exactly how the version was generated.

3.3.5 Product Quality

- Item Description: Description of the impact to product quality due to issues with computationally intensive operations (e.g., large matrix inversions, truncation and rounding). Documentation of product quality assessment (methods used, assessment summaries for each version of the datasets). Description of embedded data at the granule level including quality flags, product data uncertainty fields, data issues 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 products to ensure their proper usage. It will be helpful to document potential improvements to algorithms that (for whatever reason) were not possible to implement for the archived products.

3.3.6 Product Application

- Item Description: Useful references to published articles about the use of the data and user feedback received by the science and instrument teams about the products. Includes reports of any peculiarities or notable features observed in the products.
- Rationale: Provides additional help in understanding usage of data products besides the algorithm description and source code. History of users' assessments would be useful for understanding any issues with the products.

3.4 Mission Data Calibration

3.4.1 Calibration Method

- Item Description: The methods used for instrument/sensor radiometric and geometric calibration while in operation (e.g., in orbit). The source code used in applying the calibration algorithms. Documentation of in-line changes to calibration or to instrument or platform operations or conditions that occur throughout the mission. In the case of in situ data, station location and any changes in location, instrumentation, controlling agency, surrounding land use and other factors which could influence the long-term record.
- 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 time series ensuring consistency over time. Records of instrument operations history, environment and any mission anomalies are important to understand any quirks in data products. Source code will help users in understanding how the calibrated products are generated. It may be possible to use source code (under the right circumstances) to re-implement calibration software. All versions of software should be preserved and relation between product versions and software versions should be documented.

3.4.2 Calibration Data

- Item Description: Instrument and platform engineering data collected during operations (e.g., on orbit), including platform and instrument environment, events and maneuvers; attitude and ephemeris; aircraft position; 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. **CH01**
- 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.)

3.5 Science Data Product Software

- Item Description: Science data product generation software and software documentation. Source code used to generate 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. 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 run, including data volume and file sizes; i.e. 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 if a case can be made for future use: Description of all 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 products met requirements.
- Rationale: Product software source code and production rules provide the definitive procedural steps that document the exact implementation of the algorithm as described in algorithm theoretical basis documents. Product software information documents the relation between product versions and software versions. 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 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 product software will help users identify the source contributions to errors and uncertainties of a particular observation. 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).

3.6 Science Data Product Algorithm Inputs

- Item Description: Identify all ancillary data or other data sets used in generation or calibration of the data or derived product at all levels. Ancillary data should be stored with the 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 data set or derived product, either explicitly in data descriptions or by reference to appropriate publications. Information should include 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 granule level, include information on all inputs (including ancillary or other data granules, calibration files, look-up tables, ground control, climatology etc.) that were used to generate the product. At the appropriate level (granule or dataset) include calibration parameters, precision orbit & 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 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 versus from the instrument observations. This is especially important when investigators want to consider the impact of new improved ancillary values or ancillary geophysical relationships such as land-ocean masks or standard atmosphere profiles could impact derived climate trends, significantly reduce error or bias in a derived product. Knowledge of all algorithm inputs is critical for assessing repeatability and usability of the experiment’s results.

3.7 Science Data Product Validation

- Item Description: Datasets and documentation. Accuracy of 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 previous 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 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.

3.8 Science Data Software Tools

- Item Description: Product access (reader) tools. Software source code that would facilitate use of the calibration data, ancillary data and the data products at all levels. Includes software source code useful for creating programs that will read and display the calibration data, ancillary data and product data and 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 products.

3.9 Deliverables Checklist

- Item Description: A checklist or "meta-document" that provides documentation of how each of the content specifications discussed in the above sections are satisfied and delivered by the flight project/instrument team to the assigned ESD Data Center, to NASA STI program or will reside in a partner archive (e.g., USGS or NOAA). This checklist needs to 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 location of where some of the items are archived changes. **CH01**

- 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 ESD Data Center personnel to obtain clarifications if needed, at least as long as the points of contact are available.

This page intentionally left blank.

Abbreviations and Acronyms

CCR	Configuration Change Request
CCSDS	Consultative Committee for Space Data Systems
DAAC	Distributed Active Archive Center
DCN	Document Change Notice
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EP-TOMS	Earth Probe-Total Ozone Mapping Spectrometer
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System
ESDS	Earth Science Data System
ESIP	Federation of Earth Science Information Partners
GLAS	Geoscience Laser Altimeter System
GSFC	Goddard Space Flight Center
HIRDLS	High Resolution Dynamics Limb Sounder
ICESat	Ice, Cloud, and land Elevation Satellite
ISO	International Organization for Standardization
ITAR	International Traffic in Arms Regulations
NARA	National Archives and Records Administration
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OAIS	Open Archival Information System
RM	Reference Model
STI	Scientific and Technical Information
TBS	To be supplied
USGCRP	United States Global Change Research Program
USGS	United States Geological Survey
XFDU	XML Formatted Data Unit

This page intentionally left blank.