CEOS
Working Group on Information Systems and Services
Data Stewardship Interest Group

WGISS Data Management and Stewardship Maturity Matrix

CEOS
Data Stewardship Interest Group

Doc. Ref.: CEOS.WGISS.DSIG.DSMM
Date: October 2024
Issue: Version 1.1

Contents

Intended Audience

This document is intended to assist data managers in Earth Observation (EO) data centres with the task of ensuring Earth Observation space data sets long-term preservation, curation, accessibility, discoverability and usability.

The intended audience comprises:

1. Data providers
   - To evaluate and improve the quality and usability of their products
2. Modellers, decision-makers, and scientists
   - To improve their products
   - To make investments and take decisions
3. Data managers/stewards of data centres and repositories
   - To validate their compliance or lack of stewardship practice or standards
   - To assess the current state of their data holdings and repositories
   - To create a roadmap to improve or enhance the stewardship maturity of practices applied to all data holdings

Background and scope

This white paper presents a WGISS Data Management and Stewardship maturity assessment model in the form of a matrix for Earth Observation datasets.

In the extended environment of Maturity Matrices and Models, the Maturity Matrix for "Long-Term Scientific Data Stewardship", of Peng, and Jeffrey L. Privette, & Others (2015) [1], represents a systematic assessment model for measuring the status of individual datasets. In general, it provides information on all aspects of the data records, including all activities needed to preserve and improve the information content, quality, accessibility, and usability of data and metadata. This was used as a starting point for the WGISS Data Management and Stewardship Maturity Matrix, presented in this document.

In parallel, the GEO Data Management Principles Task Force was tasked with defining a common set of GEOSS Data Management Principles Implementation Guidelines (DMP-IG) [2]. These principles address the need for discovery, accessibility, usability, preservation, and curation of the resources made available through GEOSS.

The WGISS Data Management and Stewardship Maturity Matrix represents the result of a combined analysis on the "Long-Term Scientific Data Stewardship" Maturity Matrix and DMP-IG performed in consultation with a Data Access and Preservation Working Group at European level and with the WGISS Data Stewardship Interest Group. The resulting Matrix includes also the Research for Data Alliance (RDA) FAIR principles results [3] and the Earthnet Data Assessment Pilot (EDAP) Quality input [4] produced as part of WGISS cooperation with the CEOS Working Group on Data Calibration and Validation (WGCV).

The resulting WGISS Data Management and Stewardship Maturity Matrix fully addresses and covers data preservation, discoverability, accessibility, quality and usability aspects including data use related services and capabilities.

The WGISS Data Management and Stewardship Maturity Matrix is aligned with the following definitions:

- Data stewardship "encompasses all activities that preserve and improve the information content, accessibility, and usability of data and metadata" (National Research Council 2007);
- Data management "includes all activities for planning, execution and oversight of policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information assets." (Mosely et al. 2009).

The main objective of its application and use is to measure and verify the overall implemented or to-be-implemented data preservation and stewardship lifecycle. The Matrix can be used as a self-assessment tool to create a stewardship maturity scoreboard target for one or more scientific dataset(s), to identify actual and potential gaps, and to define a roadmap for data stewardship improvement; or to provide data quality and usability information to users, stakeholders, and decision makers. It is flexible and adaptable through a tailoring, with respect to the requirements and objectives of the data owners, highlighted after the initial process foreseen in the CEOS WGISS Data Preservation Workflow [5].

Use of the WGISS Data Management and Stewardship Maturity Matrix will allow data owners to:

- Apply a reference model for stewardship planning and resources allocation;
- Define goals and targets and create a roadmap for scientific data stewardship improvement;
- Apply detailed guidelines and recommendations for data preservation and stewardship and self-evaluate the levels and results achieved;
- Break down problems related to preservation and stewardship, and to understand the costs associated with each level;
- Provide data quality, usability information to end users, stakeholders, and decision makers.

Components and Maturity Levels of the DMSMM

The WGISS Data Management and Stewardship Maturity Matrix resulting from the above-defined work has twelve components. These components are related and mapped with the Data Management Principles Implementation Guidelines [2] and with the "Long-Term Scientific Data Stewardship" matrix [1] as follows:

WGISS Data Management & Stewardship Maturity Matrix	DMP Implementation Guidelines	Stewardship Maturity Matrix Key Components (GE-PENG)
MMP1: METADATA FOR DISCOVERY	DMP-1: Discoverability	Accessibility
MMP2: ONLINE ACCESS	DMP-2: Accessibility	Accessibility
MMP3: DATA ENCODING	DMP-3: Encoding	Usability
MMP4: DATA DOCUMENTATION	DMP-4: Documentation	Usability
MMP5: DATA TRACEABILITY	DMP-5: Traceability	Transparency /Traceability
MMP6: DATA VALIDATION	DMP-6: Quality	Assessment Data Quality Assurance
MMP7: DATA UNCERTAINTY	DMP-6: Quality	Monitoring Data Quality
MMP8: DATA QUALITY CONTROL	DMP-6: Quality	Data Quality Control
MMP9: DATA PRESERVATION	DMP-7: Preservation	Preservability
MMP10: DATA AND METADATA VERIFICATION	DMP-8: Verification	Data Integrity
MMP11: DATA PROCESSING/REPROCESSING	DMP-9: Review and reprocessing	Production Sustainability
MMP12: PERSISTENT AND RESOLVABLE IDENTIFIERS	DMP-10: Persistent and resolvable identifiers	Transparency /Traceability

A weight has been given to each WGISS Maturity Matrix component, in order to create an incremental order for their implementation. The maturity of data management and stewardship for each component, applied to individual datasets, can be assessed on a four-level measurable maturity scale:

- Level-0: Not Managed
- Level-1: Partially Managed
- Level-2: Managed
- Level-3: Fully Managed

The WGISS Data Management and Stewardship Maturity Matrix is shown in the image below.

	DISCOVERABILITY	ACCESSIBILITY	USABILITY						PRESERVATION		CURATION	
	MMP1 Metadata for Discovery	MMP2 Online Access	MMP3 Data Encoding	MMP4 Data Documentation	MMP5 Data Traceability	MMP6 Data Validation	MMP7 Metrology	MMP8 Data Quality Control	MMP9 Data Preservation	MMP10 Data Verification	MMP11 Data Processing/Reprocessing	MMP12 Persistent & Resolvable Identifier
Level-0 Not Managed	1) No catalogue available 2) No advertising available	1) Data and metadata are not accessible online	1) Data Not Structured 2) Non-standard or proprietary data format, or, poorly-documented standard file format.	1) Partial and incomplete mission documentation	1) Limited product information available (not online)	1) Validation Method - No validation activity performed. 2) Validation Results Compliance - No validation activity performed.	1) Measurement Calibration & Characterisation - Pre-flight and post-launch sensor calibration & characterisation not documented or information not available. 2) Geometric Calibration & Characterisation - Geometric calibration & characterisation not documented or not available. 3) Metrological Traceability Documentation - No traceability chain documented. 4) Uncertainty Characterisation - No uncertainty information provided. 5) Ancillary Data - Use of ancillary data undocumented.	1) No control and monitoring check 2) No quality indicator in metadata 3) No procedures documentation	1) Uncontrolled storage location. 2) Only data are stored 3) Data Records archiving not managed 4) Relevant information on Product Details Assessment not made available	1) No Data/Associated Information integrity, authenticity and readability check	1) Calibration Algorithm - Calibration algorithm not documented. 2) Geometric Processing - Geometric proccessing algorithm not documented. 3) Retrieval algorithm not documented. 4) Mission Specific Processing – Additional processing steps not documented.	1) No persistent and resolvable identifiers available
Level-1 Partially Managed	1) Advertising available 2) Catalogue search available at product level	1)Basic online services available for data and metadata access	1) Basic schema for automated data use 2) Data in documented standard file format. Non-standard naming conventions used.	1) Already existent mission documentation available and preserved for the long term 2) No link between mission documentation and data records	1) Product information available (not online)	1) Validation Method - Methodology is simple comparison, covering a limited range of satellite measurements. Uncertainty information not available for reference data. 2) Validation Results Compliance - Claimed mission performance shows some agreement with validation results.	1) Measurement Calibration & Characterisation - Pre-flight and post-launch sensor calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose. 2) Geometric Calibration & Characterisation - Geometric calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose. 3) Metrological Traceability Documentation - Traceability chain diagram and/or uncertainty tree diagram included, missing some important steps. 4) Uncertainty Characterisation - Uncertainty established by limited comparison to measurements by other sensor/s. 5) Ancillary Data- Ancillary data used in product generation, specified to some extent, though incomplete. Not entirely of a sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.	1) Basic data quality control and monitoring check 2) Minimal set of quality control procedures documented and available	1) Basic archiving for original data records preservation 2) Assessment of SW preservation 3) Product Details Assessment: Any required information missing	1) Data Records/Associated Information integrity basic check	1) Calibration Algorithm - Calibration algorithm somewhat documented. Calibration algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. 2) Geometric Processing - Geometric processing documented. Missing all or part of the calibration parameters. Calibration algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. Confidence in the calibration quality is minimal. 3) Retrieval algorithm somewhat documented. Retrieval algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. 4) Mission Specific Processing – Additional processing steps documented. Additional processing steps not considered fit for stated purpose.	1) Persistent identifier assignment only for particular Data Records Collections 2) Basic landing pages management
Level-2 Managed	1) Detailed catalogue search available at product level 2) Product metadata oriented towards an international standard 3) Data Collection and Associated Information searchable. 4) International standard for Collection metadata	1) Simple Access Architecture through metadata 2) Data access system oriented towards an international standard	1) Use of non-proprietary international standards encodings for syntactic interoperability. 2) Periodically repackaging/reformatting of archived data. 3) Data in well-documented standard file format, community naming convention standards.	1) Documentation produced, published and well described 2) Link between mission documentation and data records created and managed	1) Dataset tested for presence of correct provenance metadata. 2) Well described product information available online	1) Validation Method - Methodology assesses satellite measurements and reference data with respect to their characterised uncertainties. Reference measurements are assessed to be well representative of the satellite measurements. 2) Validation Results Compliance- Claimed mission performance shows excellent agreement with validation results. Analysis performed independently of satellite mission owner	1) Measurement Calibration & Characterisation - Pre-flight and post-launch measurement calibration & characterisation efforts cover all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Pre-flight calibration is traceable to SI or standard reference, characterisation methods meet good practice. Post-launch Cal/Val uses appropriate community infrastructure/methods (e.g. RadCalNet). 2)Geometric Calibration & Characterisations - Geometric calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Post-launch characterisation uses appropriate community infrastructure/methods. 3)Metrological Traceability Documentation - Rigorous uncertainty tree diagram, with a traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. 4) Uncertainty Characterisation - Full GUM approach to estimate measurement uncertainty, all important sources of uncertainty included. Uncertainty per pixel provided. 5) Ancillary Data - Ancillary data used in product generation, fully specified per product, and traceable. Ancillary data used are of sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.	1) Quality indicator post-processing available 2) Quality control procedures documented and available online	1) Preservation repository certified internally 2) Community-standard for archiving metadata 3) Product Details Assessment: All required information available, any recommended information missing	1) Data Records/Associated Information content integrity check and verification 2) Media readability and accessibility testing	1) Calibration Algorithm - Calibration algorithm documented. Calibration used "fit for purpose" in terms of the mission's stated performance all expected use cases. 2) Geometric Processing - Geometric processing documented. All input calibration parameters exist. Methodology used is considered "fit for purpose" in terms of the mission's stated performance for all expected use cases. Quality flags indicate good geometric accuracy with less than 5% exceptional. 3) Retrieval algorithm documented. Retrieval algorithm "fit for purpose" in terms of the mission's stated performance all expected use cases and validated performance against similar algorithms or with empirical evidence. 4) Mission Specific Processing – Additional processing steps documented. All additional processes steps considered fit for stated purpose.	1) Persistent identifier assignment to all disseminated Data Records Collections and metadata 2) Automatic landing page generation and extensive management of landing pages
Level-3 Fully Managed	1) Catalogue accessible via international or community agreed standards protocol 2) Data policy available in metadata 3) Periodic updates of metadata in the catalogue 4) Quality indicator metadata available and discoverable 5) Search results relevancy. 6) Seamless transition from discovery to access	1) International standard for Data and metadata access system 2) Data policy available in the metadata. 3) Visualisation services 4) Reporting system 5) Hosted processing 6) Quick adoption to new technologies and standards evolution 7) Data and metadata accessible through a free and open access protocol	1) Accepted and Available semantic encoding standards for complete interoperability. 2) Data and metadata uses FAIR-compliant vocabularies 3) Analysis Ready Data standard	1) Standards based metadata for documentation 2) Link between mission documentation and data records published	1) Automatic metadata generation for provenance documentation 2) Complete and updated data provenance available online	1) Validation Method - Methodology assesses satellite measurements and reference data with respect to their error-covariance and attempts to validate those uncertainties. Reference measurements independently assessed to be fully representative of the satellite measurements. 2) Validation Results Compliance - Claimed mission performance shows excellent agreement with validation results, measurement uncertainties also validated. Analysis performed independently of satellite mission owner.	1) Measurement Calibration & Characterisation - In addition to meeting Excellent criteria, calibration and characterisation include the measurements needed to assess uncertainties at the component level and their impact on the final product. Post-launch Cal/Val uses appropriate community infrastructure/methods traceable to SI (e.g. FRMs, RadCalNet). 2)Geometric Calibration & Characterisations - As for Level-2, In addition geometric calibration and characterisation includes the measurements needed to assess uncertainties at the component level and their impact on the final product. The quality is "fit for purpose" in terms of the mission's stated performance, and meets the science users expectations. 3)Metrological Traceability Documentation - Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI. 4)Uncertainty Characterisation - Full GUM approach is used to estimate measurement uncertainty, including a treatment of error-covariance. Per pixel uncertainties in components, e.g., random systematic – as appropriate for the error-correlation structure of the data. 5)Ancillary Data - As Level-2, ancillary data traceable to SI where appropriate	1) Data quality control fully compliant with an international standard 2) Quality indicator pre and post processing available in the metadata 3) Quality metadata assessed	1) Preservation repository officially certified 2) Periodic technology refreshment 3) Identify and manage the basic preservation of relevant mission SW, ensuring that preserved data can be recreated. 4) Continuity of service availability 5) Product Details Assessment: All required and recommended information available	1) Automatic Data Records/Associated Information content integrity check and verification 2) Data authenticity verifiable internally and by the final user 3) Automatic verification process, including monitoring and reporting	1) Calibration Algorithm - Calibration algorithm well-documented. State-of-the-art calibration algorithm applied and considered "fit for purpose" in terms of the mission's stated performance. 2) Geometric Processing - Geometric processing well-documented. State-of-the-art methodology used, easily "fit for purpose" in terms of the mission's stated performance. Quality flags indicate excellent geometric accuracy. 3) Retrieval algorithm documented. State-of-the-art retrieval "fit for purpose" in terms of the mission's stated performance, full uncertainty budget derived and validated. 4) Mission Specific Processing – Additional processing steps documented. All additional processes steps considered fit for stated purpose.	1) Persistent identifier created for all accessible data records and metadata 2) Metadata includes the identifier for the data 3) Metadata is offered in such a way that it can be harvested and indexed

Earth observation domain applicability

The WGISS Data Management and Stewardship Maturity Matrix may be adopted to facilitate and improve CEOS Agencies Data Management and Stewardship activities and achievements. It can be further tailored to take into account specific Earth Observation requirements and pre-existing Best Practices available internationally or at Agency level.

A dataset appraisal activity should initially define the desired level to be reached for each maturity matrix component, for a specific mission/dataset, for example based on:

- Mapping versus final user exploitation capabilities;
- Mapping w.r.t. data preservation commitments, budgets, responsibilities and preservation requirements.

Different missions' datasets can have different targets and varying maturity level ratings. More details on the different maturity matrix components and levels are provided in the sections below.

WGISS Maturity Matrix components and associated maturity levels

DOMAIN: DISCOVERABILITY

MMP-1: METADATA FOR DISCOVERY

Data and all associated metadata will be discoverable through catalogues and search engines. Data access and use conditions, including licenses, will be clearly indicated.

Level-0	1) No catalogue available. 2) No advertising available.
Level-1	1) Advertising available. 2) Catalogue search available at product level.
Level-2	1) Detailed catalogue search available at product level. 2) Product metadata oriented towards an international standard. 3) Data Collection and Associated Information searchable. 4) International standard for Collection metadata.
Level-3	1) Catalogue accessible via international or community agreed standards protocol. 2) Data policy available in metadata. 3) Periodic updates of metadata in the catalogue. 4) Quality indicator metadata available and discoverable. 5) Search results relevancy. 6) Seamless transition from discovery to access.

DOMAIN: ACCESSIBILITY

MMP-2: ONLINE ACCESS

Data will be accessible via online services, including, at a minimum, direct download, but preferably user customisable services for access, visualisation and analysis.

Level-0	1) Data and metadata not accessible online.
Level-1	2) Basic online services available for data and metadata access.
Level-2	1) Simple Access Architecture through metadata. 2) Data access system oriented towards an international standard.
Level-3	1) International standard for Data and metadata access system. 2) Data policy available in the metadata. 3) Visualisation services. 4) Reporting system. 5) Hosted processing. 6) Quick adoption to new technologies and standards evolution. 7) Data and metadata accessible through a free and open access protocol.

DOMAIN: USABILITY

MMP-3: DATA ENCODING

Data should be structured using encodings that are widely accepted in the target user community and aligned with organisational needs and observing methods, with preference given to non-proprietary international standards.

Level-0	1) Data Not Structured. 2) Non-standard or proprietary data format, or, poorly-documented standard file format.
Level-1	1) Basic schema for automated data use. 2) Data in documented standard file format; Non-standard naming conventions used.
Level-2	1) Use of non-proprietary international standards encodings for syntactic interoperability. 2) Periodically repackaging/reformatting of archived data. 3) Data in well-documented standard file format, community naming convention standards.
Level-3	1) Accepted and available semantic encoding standards for complete interoperability. 2) Data and metadata use FAIR-compliant vocabularies. 3) Analysis Ready Data standard.

MMP-4: DATA DOCUMENTATION

Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata, based on international or community approved standards. To the greatest extent possible, data will also be described in peer-reviewed publications and referenced in the metadata record.

Level-0	1) Partial and incomplete mission documentation.
Level-1	1) Already existent mission documentation available and preserved for the long term. 2) No link between mission documentation and data records.
Level-2	1) Documentation produced, published and well described. 2) Link between mission documentation and data records created and managed.
Level-3	1) Standards based metadata for documentation. 2) Link between mission documentation and data records published.

MMP-5: DATA TRACEABILITY

Data will include provenance metadata, indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain.

Level-0	1) Limited product information available (not online).
Level-1	1) Product information available (not online).

Level-2	1) Dataset tested for presence of correct provenance metadata. 2) Well described product information available online.
Level-3	1) Automatic metadata generation for provenance documentation. 2) Complete and updated data provenance available online.

MMP-6: DATA VALIDATION

Data will be validated and assessed in terms of consistency between both the data values and their uncertainties with those of independent reference data.

Level-0	1) Validation Method - No validation activity performed. 2) Validation Results Compliance - No validation activity performed.
Level-1	1) Validation Method - Methodology is simple comparison, covering a limited range of satellite measurements. Uncertainty information not available for reference data. 2) Validation Results Compliance - Claimed mission performance shows some agreement with validation results.
Level-2	1) Validation Method - Methodology assesses satellite measurements and reference data with respect to their characterised uncertainties. Reference measurements are assessed to be well representative of the satellite measurements. 2) Validation Results Compliance - Claimed mission performance shows excellent agreement with validation results. Analysis performed independently of satellite mission owner.
Level-3	1) Validation Method - Methodology assesses satellite measurements and reference data with respect to their error-covariance and attempts to validate those uncertainties. Reference measurements independently assessed to be fully representative of the satellite measurements. 2) Validation Results Compliance - Claimed mission performance shows excellent agreement with validation results, measurement uncertainties also validated. Analysis performed independently of satellite mission owner.

MMP-7: DATA METROLOGY

To ensure measurements are both meaningful and defensible it is crucial that they come with rigorously evaluated uncertainty estimates. This section of the mission quality assessment evaluates the methodology used to estimate uncertainty values for a given mission, the extent of the mission's analysis and how the values are provided.

Level-0	1) Measurement Calibration & Characterisation - Pre-flight and post-launch sensor calibration & characterisation not documented or information not available. 2) Geometric Calibration & Characterisation - Geometric calibration & characterisation not documented or not available. 3) Metrological Traceability Documentation - No traceability chain documented. 4) Uncertainty Characterisation - No uncertainty information provided. 5) Ancillary Data - Use of ancillary data undocumented.

Level-1	1) Measurement Calibration & Characterisation - Pre-flight and post-launch sensor calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a sufficient level of quality to be judged fit for purpose. 2) Geometric Calibration & Characterisation - Geometric calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a sufficient level of quality to be judged fit for purpose. 3) Metrological Traceability Documentation - Traceability chain diagram and/or uncertainty tree diagram included, missing some important steps. 4) Uncertainty Characterisation - Uncertainty established by limited comparison to measurements by other sensor(s). 5) Ancillary Data - Ancillary data used in product generation, specified to some extent, though incomplete. Not entirely of a sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.
Level-2	1) Measurement Calibration & Characterisation - Pre-flight and post-launch measurement calibration & characterisation efforts cover all reasonable aspects of instrument behaviour to a quality that it is "fit for purpose" in terms of the mission's stated performance. Pre-flight calibration is traceable to SI or standard reference, characterisation methods meet good practice. Post-launch Cal/Val uses appropriate community infrastructure/methods (e.g. RadCalNet). 2) Geometric Calibration & Characterisations - Geometric calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Post-launch characterisation uses appropriate community infrastructure/methods. 3) Metrological Traceability Documentation - Rigorous uncertainty tree diagram, with a traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. 4) Uncertainty Characterisation - Full GUM approach to estimate measurement uncertainty, all important sources of uncertainty included. Uncertainty per pixel provided. 5) Ancillary Data - Ancillary data used in product generation, fully specified per product, and traceable. Ancillary data used are of sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.
Level-3	1) Measurement Calibration & Characterisation - In addition to meeting Excellent criteria, calibration and characterisation include the measurements needed to assess uncertainties at the component level and their impact on the final product. Post-launch Cal/Val uses appropriate community infrastructure/methods traceable to SI (e.g. FRMs, RadCalNet). 2) Geometric Calibration & Characterisations - As for Level-2, in addition geometric calibration and characterisation includes the measurements needed to assess uncertainties at the component level and their impact on the final product. The quality is "fit for purpose" in

	terms of the mission's stated performance and meets the science users expectations. 3) Metrological Traceability Documentation - Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI. 4) Uncertainty Characterisation - Full GUM approach is used to estimate measurement uncertainty, including a treatment of error-covariance. Per pixel uncertainties in components, e.g., random systematic – as appropriate for the error-correlation structure of the data. 5) Ancillary Data - As Level-2, ancillary data traceable to SI where appropriate.

MMP-8: DATA QUALITY-CONTROL

Data will be quality controlled and the results of quality control shall be indicated in the metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

Level-0	1) No control and monitoring check. 2) No quality indicator in metadata. 3) No procedures documentation.
Level-1	1) Basic data quality control and monitoring check. 2) Minimal set of quality control procedures documented and available.
Level-2	1) Quality indicator post-processing available. 2) Quality control procedures documented and available online.
Level-3	1) Data quality control fully compliant with an international standard. 2) Quality indicator pre and post processing available in the metadata. 3) Quality metadata assessed.

DOMAIN: PRESERVATION

MMP-9: DATA PRESERVATION

Data will be protected from loss and preserved for future use; preservation planning will be for the long term, and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.

Level-0	1) Uncontrolled storage location. 2) Only data are stored. 3) Data Records archiving not managed. 4) Relevant information on Product Details Assessment not made available.
Level-1	1) Basic archiving for original data records preservation. 2) Assessment of SW preservation. 3) Product Details Assessment: Any required information missing.

Level-2	1) Preservation repository certified internally: - Documented storage procedures (planning of periodic media refreshment). - Redundancy managed (e.g. back-up, different media technology). - Basic archiving processes measured and controlled. 2) Community-standard for archiving metadata (e.g. AIP). 3) Product Details Assessment: All required information available, any recommended information missing.
Level-3	1) Preservation repository officially certified (e.g. ISO 19363, CoreTrustSeal). 2) Periodic technology refreshment. 3) Identify and manage the basic preservation of relevant mission SW, ensuring that preserved data can be recreated. 4) Continuity of service availability (Business Continuity, Disaster and Recovery, etc.). 5) Product Details Assessment: All required and recommended information available.

MMP-10: DATA AND METADATA VERIFICATION

Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.

Level-0	1) No Data/Associated Information integrity, authenticity and readability check.
Level-1	1) Data Records/Associated Information integrity basic check (e.g. checksum).
Level-2	1) Data Records/Associated Information content integrity check and verification. 2) Media readability and accessibility testing.
Level-3	1) Automatic Data Records/Associated Information content integrity check and verification. 2) Data authenticity verifiable internally and by the final user. 3) Automatic verification process, including monitoring and reporting.

DOMAIN: CURATION

MMP-11: DATA PROCESSING/REPROCESSING

Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate; where applicable, this shall follow established and agreed procedures.

Level-0	1) Calibration Algorithm - Calibration algorithm not documented. 2) Geometric Processing - Geometric processing algorithm not documented. 3) Retrieval algorithm not documented. 4) Mission Specific Processing - Additional processing steps not documented.

Level-1	1) Calibration Algorithm - Calibration algorithm somewhat documented. Calibration algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. 2) Geometric Processing - Geometric processing documented. Missing all or part of the calibration parameters. Calibration algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. Confidence in the calibration quality is minimal. 3) Retrieval algorithm somewhat documented. Retrieval algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. 4) Mission Specific Processing - Additional processing steps documented. Additional processing steps not considered fit for stated purpose.
Level-2	1) Calibration Algorithm - Calibration algorithm documented. Calibration used "fit for purpose" in terms of the mission's stated performance for all expected use cases. 2) Geometric Processing - Geometric processing documented. All input calibration parameters exist. Methodology used is considered "fit for purpose" in terms of the mission's stated performance for all expected use cases. Quality flags indicate good geometric accuracy with less than 5% exceptional. 3) Retrieval algorithm documented. Retrieval algorithm "fit for purpose" in terms of the mission's stated performance for all expected use cases and performance validated against similar algorithms or with empirical evidence. 4) Mission Specific Processing - Additional processing steps documented. All additional processes steps considered fit for stated purpose.
Level-3	1) Calibration Algorithm - Calibration algorithm well-documented. State-of-the-art calibration algorithm applied and considered "fit for purpose" in terms of the mission's stated performance. 2) Geometric Processing - Geometric processing well-documented. State-of-the-art methodology used, easily "fit for purpose" in terms of the mission's stated performance. Quality flags indicate excellent geometric accuracy. 3) Retrieval algorithm documented. State-of-the-art retrieval "fit for purpose" in terms of the mission's stated performance, full uncertainty budget derived and validated. 4) Mission Specific Processing - Additional processing steps documented. All additional processes steps considered fit for stated purpose.

MMP-12: PERSISTENT AND RESOLVABLE IDENTIFIERS

Data will be assigned appropriate persistent, unique and resolvable identifiers to enable documents to cite the data used and to enable data providers to receive Information for the use of their data.

Level-0	1) No persistent and resolvable identifiers available.
Level-1	1) Persistent identifier assignment only for particular Data Records Collections. 2) Basic landing pages management (e.g. manual generation and updates, no common template).
Level-2	1) Persistent identifier assignment to all disseminated Data Records Collections and metadata. 2) Automatic landing page generation and extensive management of landing pages.
Level-3	1) Persistent identifier [6] created for all accessible data records and metadata. 2) Metadata includes the identifier for the data. 3) Metadata is offered in such a way that it can be harvested and indexed.

DMSMM User Guide

Getting started

- Download the latest DSMM template file from https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix_Schema%20to%20be%20filled.xlsx
- Rename the template file, using the following file naming convention: DMSMM<DatasetName>_<Version>_<Date><Point of Contact>
- Go through each key component in the template, identify the stewardship practices applied to the dataset, and document your rating and justifications, including any relevant URLs. Add comment if necessary. Use dark brown text to denote uncertainty about the practice or the fact that additional information or research is needed.
- All stewardship practices applied to the dataset, not just ones by the repository/archive, should be captured.
- It is recommended to identify the community in the justification section that stewardship practices and standards are pertained to, based on the categories of communities defined in the template or defined in your own words.
- Obtain any additional information, if necessary, from product website(s), in the literature, and/or from POCs of dataset, archive, and access.
- Review the results and fill in the matrix cells with the defined color scheme.
- Capture the assessment results in the DSMM metadata section
- The following presentation gives more details on the use of the DMSMM: https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix_User%20Guide.pdf

.

Reference Document

[1] Peng, Privette, & Others, **Scientific Data Stewardship Maturity Matrix**
1. https://datascience.codata.org/articles/abstract/10.2481/dsj.14-049/
2. http://tinyurl.com/DSMMtemplate
3. http://www.slideshare.net/gepeng86/scientific-data-stewardship-maturity-matrix
4. http://tinyurl.com/DSMM-FlowChart

[2] GEOSS, "**Data Management Principles**", https://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf

[3] Research Data Alliance FAIR data maturity model Working Group "**FAIR Data Maturity Model: specification and guidelines**", https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines

[4] EDAP "**Quality Assessment Guidelines**", https://earth.esa.int/web/sppa/activities/edap-best-practice-guidelines

[5] CEOS, "**Long Term Preservation of Earth Observation Space Data: Preservation Workflow**", http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf

[6] CEOS, "**CEOS Persistent Identifier Best Practices**", https://ceos.org/document_management/Working_Groups/WGISS/Documents/WGISS%20Best%20Practices/CEOS%20Persistent%20Identifier%20Best%20Practices.pdf

Related Documentation

[7] CEOS, "**EO Data Preservation Guidelines Best Practices**", https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Data%20Preservation%20Guidelines.pdf

[8] CEOS, "**EO Preserved Data Set Content**", http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf

[9] CEOS, "**Technical Content and Associated Information Preservation Best Practices**",

https://ceos.org/document_management/Working_Groups/WGISS/Documents/WGISS%20Best%20Practices/CEOS%20Technical%20Content%20and%20Associated%20Information%20Preservation%20Best%20Practices.pdf

[10] CEOS, "**Generic Earth Observation Data Set Consolidation Process**", http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/GenericEarthObservationDataSetConsolidationProcess_v1.0.pdf

[11] CEOS, "**Long-Term Preservation of Earth Observation Space Data: Glossary of Acronyms and Terms**", https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf