

CEOS WGISS Data Management and Stewardship Maturity Matrix and Application at ESA

Mirko Albani¹, Iolanda Maggio²

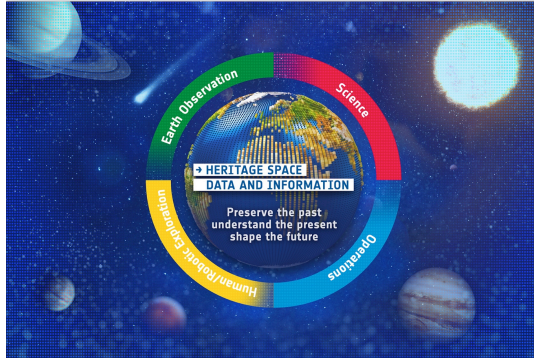
1) European Space Agency (ESA) ESRIN

2) RHEA GROUP c/o ESA ESRIN

Pre-ESIP Workshop

13 July 2020

CEOS WGISS Data Stewardship Best Practices



Chair, Robert Woodcock, CSIRO

Vice-Chair, Makoto Natsuisaka, JAXA

WGISS Principals

CAS/AOE: Guangyu Liu
CAS/RADI: Lizhe Wang
CSA: Paul Brand
CNES: Richard Moreno
CONAE: Homero Lloza
CSIRO: Robert Woodcock
DLR: Karin Mädicke
ESA: Mirko Albani
EUMETSAT: Michael Schick
GA: Simon Oliver
HSC: Gabor Remetey
INPE: Lúbia Vinhas
ISRO: Nitant Dubé
JAXA: Makoto Natsuisaka
NASA: Andrew Mitchell
NOAA: Kenneth Casey
NRSCC: Chuang Liu
ROSCOSMOS: Tamara Ganina
USGS: Thomas Solbre
UKSA: Esther Conway

WGISS Secretariat (Michelle Piegras, CSIRO)

Data Preservation and Stewardship Interest Group (Mirko Albani, ESA)

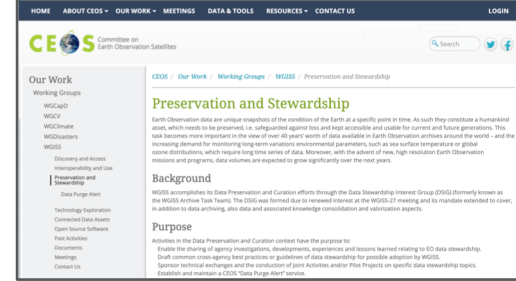
Discovery and Access - WGISS Connected Data Assets System Level Team (Damiano Guerucci, ESA)

- IDN (Michael Morahan, NASA)
- FedEO (Mirko Albani, ESA)
- CWIC (Minnie Wong, NASA)
- OpenSearch II (Olivier Barois, ESA)

Interoperability and Use Interest Group (Robert Woodcock, CSIRO; Richard Moreno, CNES)

- Future Data Architectures (Robert Woodcock, CSIRO)
- Carbon Portal (Richard Baldwin, NOAA; Lijng Di, NOAA)
- Data Analytics Software and Tools (Mirko Albani, ESA)

Technology Exploration Interest Group (Chris Lynnes, NASA; Yousuke Ishihata, JAXA)



ESA UNCLASSIFIED - For Official Use

ESA | 13/07/2020 | Slide 2



European Space Agency

What is the Maturity Matrix/Model?



Who could use it?

Why should it be used?

DMSMM defines all activities needed to preserve and improve the information content, quality, accessibility, and usability of data and metadata.

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).

CEOS WGISS DMSMM: Generation process

MM for Long-Term Scientific Data Stewardship

GEOSS Data Management Principles

RDA FAIR Data Maturity Model

WMO Stewardship MM for Climate Data

ESA Earthnet Quality Data Assessment Pilot

Discoverability	Accessibility	Usability	Preservation	Open Quality	Open Quality	Open Quality	Transparency	Other
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata
Level 1: Metadata	Level 2: Metadata	Level 3: Metadata	Level 4: Metadata	Level 5: Metadata	Level 6: Metadata	Level 7: Metadata	Level 8: Metadata	Level 9: Metadata

DISCOVERABILITY DMP-1: Data and metadata will be discoverable ACCESSIBILITY DMP-2: Data will be accessible via online services USABILITY DMP-3: Encoding DMP-5: Traceability PRESERVATION DMP-7: Preservation CURATION DMP-9: Review and reprocessing DMP-10: Persistent and resolvable identifiers	DMP-4: Documentation DMP-6: Quality DMP-8: Verification
--	--



Data Access	Usability & Usage	Quality Management	Data Management
Discoverability	Data Portability	Quality Assurance & Control	Preservation
Accessibility	Documentation Usage	Quality Assessment	Metadata Governance
		Data Integrity	
The state or ability to locate (Discoverability) and get to the dataset (Accessibility)	How easily the data product may be understood and integrated by users; the usage and impact of the dataset	The state of quality assurance, control, and assessment; data security and reliability, and data free from corruption	The state of the dataset preservation, metadata completeness, and governance practices

Product Information	Product Generation	Product Information	Uncertainty Characterisation	Validation
Product Details	Sensor Calibration & Characterisation Pre-Flight	Product Flags	Uncertainty Characterisation Method	Reference Data Representativeness
Availability & Accessibility	Sensor Calibration & Characterisation Post-Flight	Additional Information	Uncertainty Sources Included	Reference Data Quality
Product Format	Additional Processing		Uncertainty Values Provided	Validation Method
User Documentation			Geolocation Uncertainty	Validation Results
Methodological Traceability Documentation				

Review Cycle



Review Cycle



CEOS WGISS Data Management & Stewardship Maturity Matrix



	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 Data Uncertainty	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	Data and metadata are not accessible online	1) Data Not Structured 2) Non-standard or proprietary data format, or poorly-documented standard file format.	Partial and incomplete mission documentation	Limited product information available (not online)	1) Reference Data Representativeness: No validation 2) Reference Data Quality: No validation 3) Validation Method - No validation 4) Validation Results - No validation 5) Uncertainty Values: No uncertainty information provided.	1) Uncertainty Method: Uncertainty characterisation not performed, or method not documented. 2) Uncertainty Sources: Uncertainty characterisation not performed, or sources analysed not documented. 3) Uncertainty Values: No uncertainty information provided.	1) No control and monitoring check 2) No quality indicator in metadata 3) No procedures documented	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	No Data/Associated Information integrity, authenticity and readability check	1) No reprocessing activities planned 2) Pre-flight calibration & characterisation not documented or information not available. 3) Post-launch calibration & characterisation not documented or not available 4) Processing: Additional processing steps not documented.	No persistent and resolvable identifiers available
Level-1 Partially Managed	1) Advertising available 2) Catalogue search available at product level with minimum set of metadata	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 assessed to be mostly representative of the satellite measurements, covering a primary range of satellite measurements and at ad hoc opportunities 2) Reference Data Quality: single uncertainty for the entire dataset. 3) Validation Method: simple uncertainty estimated 4) Validation Results: Validation results show good agreement between satellite and reference measurements within uncertainties in most cases.	Product information available (not online)	1) Reference Data Representativeness: measurements assessed to be mostly representative of the satellite measurements, covering a primary range of satellite measurements and at ad hoc opportunities 2) Reference Data Quality: single uncertainty for the entire dataset. 3) Validation Method: simple uncertainty estimated 4) Validation Results: Validation results show good agreement between satellite and reference measurements within uncertainties in most cases.	1) Uncertainty Method: Limited use of GUM approach, and/or, an expanded comparison to measurements by other sensors. 2) Uncertainty Sources: Most important sources of uncertainty included. 3) Uncertainty Values: Single uncertainty value provided for subsets of data	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	Data Records/Associated Information integrity basic check	1) Minor updates and bugs corrections of data records implemented 2) Data Records repackaging and/or reformatting 3) Pre-flight 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 4) Post-launch 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. 5) Additional processing steps documented. Some important additional processing steps may not be 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 Records Collection and Associated Information searchable. 4) Collection metadata oriented towards an international standard	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 archive 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	Dataset tested for presence of correct provenance metadata. Well described product information available online	1) Reference Data Representativeness: measurements assessed to be well representative of the satellite measurements, covering a reasonable range of the satellite measurements and carried out using FRM or community approved methods. 2) Reference Data Quality: full uncertainty information, assessed following the GUM and traceable to community reference or SI 3) Validation Methods assess satellite measurements and reference data w.r.t. their uncertainties 4) Validation Results show excellent agreement between satellite and reference measurements, within uncertainties. Analysis performed independently of satellite mission owner.	1) Uncertainty Method: GUM approach to estimate measurement uncertainty with full breakdown of components and separated as Type A or B classification. 2) Uncertainty Sources: All important sources of uncertainty included 3) Uncertainty Values: Total uncertainty per pixel is provided, with basic breakdown of key components, no error-covariance.	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) Reprocessing for calibration and/or algorithm improvement 2) Pre-flight 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. Calibration traceable to SI or community reference, characterisation meets good practice. 3) Post-launch 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 and uses appropriate community infrastructure/methods (CEOS/FRMs). 4) Processing: Additional processing steps documented. All additional processes steps 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) Product rich metadata fully compliant with an international standard 2) Collection metadata fully compliant with an international standard 3) Catalogue accessible via an accepted international or community agreed upon standards protocol 4) Data policy on the use conditions/restrictions and legal constraints of the data, available in metadata 5) Periodic updates of metadata in the catalogue 6) Quality indicator metadata available and discoverable 7) Search results ordered by relevancy 8) Seamless transition from discovery to access	1) Data and metadata access system fully compliant with an international standard 2) Data policy regarding use conditions and restrictions of the data, available in the metadata. 3) Visualisation services allowing a user to view images of data 4) Reporting system available 5) Hosted processing 6) Quick adoption to new technologies and standards evolution 7) Data and metadata accessible through a free 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) Reference Data Representativeness: Reference measurements independently assessed to be fully representative of the satellite measurements, covering the satellite's full range of measurements and with full assessment of uncertainties and carried out on a regular basis using product performance. 2) Reference Data Quality: full uncertainty and error-correlation information, assessed following the GUM and traceable to SI 3) Validation Methods assess satellite measurements and reference data w.r.t. their error-covariance and validates those uncertainties. 4) Validation Results show excellent agreement between satellite and reference measurements, within uncertainties. Uncertainty validated. Analysis performed independently of satellite mission owner.	1) Uncertainty Method: GUM approach to estimate measurement uncertainty, including a treatment of error-covariance. 2) Uncertainty Sources: All reasonable sources of uncertainty included. 3) Uncertainty Values: Uncertainties per pixel provided with error-covariance information for all appropriate components.	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 internally and by the final user 2) Data authenticity verifiable internally and by the final user 3) Automatic verification process, including monitoring and reporting	1) Reprocessing for time-series creation 2) Roadmap for technology evolution 3) Fidelity of accurate and relevant attributes are provided to allow reuse 4) Metadata includes information about the licence under which the data can be reused 5) Pre-flight: As Level-2, additionally calibration and characterisation includes the measurements needed to assess uncertainties at component level and their impact on the final product. 6) Post-launch 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. Measurements fully traceable to SI or community reference at an uncertainty commensurate with the product specification and carried out regularly across the full range of observational conditions of the product and dynamic range. 7) Processing: All additional processing steps fully documented and state-of-the-art.	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

http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf

CEOS WGISS Data Management & Stewardship

Maturity Matrix application at ESA



❑ An instrument to:

- Fix objectives and targets for space data holdings management and stewardship (taking into account applicable data policies and budget constraints)
- Monitor progress and achievements and measure effectiveness of the organization applied stewardship processes

❑ Applicable at mission (initial planning) and dataset (verification stage) level and tailored case by case

❑ Currently used for ESA EO heritage missions and datasets holdings (plan to extend use to operational missions)

❑ Periodic assessment to measure gaps in the implemented processes wrt Data stewardship objectives

CEOS WGISS Data Management & Stewardship

Maturity Matrix application at ESA - AVHRR Example



	Discoverability	Accessibility	Usability				Preservation		Curation	
	DMP-1 Metadata for Discovery	DMP-2 Online Access	DMP-3 Data Encoding	DMP-4 Data Documentation	DMP-5 Traceability	DMP-6 Quality	DMP-7 Preservation	DMP-8 Verification	DMP-9 Reprocessing	DMP-10 Persistent Identifier
Level-0 Not Managed	1) No catalogue available 2) No advertising available	Data are not accessible online	Data Not Structured	Partial and incomplete mission documentation (mission, data and product documentation)	Limited product information available (not online)	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	No Data/Associated Knowledge integrity, authenticity and readability check	No reprocessing activities planned	No persistent and resolvable identifiers available
Level-1 Partially Managed	1) Advertising available 2) Catalogue search available at product level with minimum set of metadata	Basic online services available for data access (e.g. FTP/HTTP direct download)	Basic schema for automated data use	1) Already existent mission documentation available and preserved for the long term 2) No link between mission documentation and data records	Product information available (not online)	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: - The entity in charge of data long term preservation is identified and designated - Minimal redundancy and metadata preservation 2) Assessment of SW preservation	Data Records/Associated Knowledge integrity basic check (e.g. checksum)	1) Minor updates and bugs corrections of data records implemented 2) Data Records repackaging and/or reformatting	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 Managed	1) Detailed catalogue search available at product level 2) Product metadata oriented towards an international standard (e.g. ISO, OGC, INSPIRE, etc.) 3) Data Records Collection and Associated Knowledge searchable. 4) Collection metadata oriented towards an international standard (e.g. ISO, OGC, INSPIRE, etc.)	1) Simple Access Architecture through metadata - e.g. Data Access through a catalogue service 2) Data access system oriented towards an international standard (e.g. OpenSearch, ISO)	1) Use of non-proprietary international standards encodings for syntactic interoperability. If a proprietary format is used, it has to be formally and semantically described. 2) Periodically repackaging/ reformatting of archived data	1) Documentation produced, published and well described (covering the format, metadata, and methods used in creating and validating the data) 2) Link between mission documentation and data records created and managed (internal use only)	Dataset tested for presence of correct provenance metadata (presence, completeness and correctness). Well described product information available online	1) Quality indicator post-processing available 2) Quality control procedures documented and available online	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)	1) Data Records/Associated Knowledge content integrity check and verification 2) Media readability and accessibility testing	Reprocessing for calibration and/or algorithm improvement	1) Persistent identifier assignment to all disseminated Data Records Collections 2) Automatic landing page generation and extensive management of landing pages
Level-3 Well Managed	1) Product metadata fully compliant with an international standard (e.g. ISO, OGC, INSPIRE, etc.) 2) Collection metadata fully compliant with an international standard (e.g. ISO, OGC, INSPIRE, etc.) 3) Catalogue accessible via an accepted international or community agreed upon standards protocol 4) Data policy on the use conditions/restrictions and legal constraints of the data, available in metadata 5) Periodic updates of metadata in the catalogue (e.g. contact point) 6) Quality indicator metadata available and discoverable 7) Search results ordered by relevancy. 8) Seamless transition from discovery to access	1) Data access system fully compliant with an international standard (e.g. OpenSearch, ISO) 2) Data policy regarding use conditions and restrictions of the data, available in the metadata. 3) Visualisation services allowing a user to view images of data (e.g. Web Map Services for geospatial data, browse image services) 4) Reporting system available (e.g. user statistics, data access reports, system availability reports, etc.) 5) Hosted processing (e.g. on the fly processing) 6) Quick adoption to new technologies and standards evolution	Accepted and Available semantic encoding standards for complete interoperability	1) Standards based metadata for documentation (e.g. to support the reproducibility of science) 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) 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 (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.)	1) Automatic Data Records/Associated Knowledge content integrity check and verification 2) Data authenticity verifiable internally and by the final user 3) Automatic verification process, including monitoring and reporting	1) Reprocessing for time-series creation (e.g. FDR for ECV) 2) Roadmap for technology evolution	Persistent identifier created for all accessible data records

Next Steps (for discussion)

How to assign levels?

Example for DMP-6 Validation

Reference Data

Representativeness = a

Reference Data Quality = b

Validation Method = c

Validation Results = d

a {0,1,2,3}

b {0,1,2,3}

c {0,1,2,3}

d {0,1,2,3}

Values assigned to each parameter according to level of maturity

Level0: abcd = 0

Level1: 1 abcd 40

Level2: 41 abcd 80

Level3: abcd = 3x3x3x3=81

	DMP-6 Validation NEW
<ul style="list-style-type: none"> Level-0 Not Managed 	<ul style="list-style-type: none"> Reference Data Representativeness - No validation activity performed. Reference Data Quality - No validation activity performed. Validation Method - No validation activity performed. Validation Results - No validation activity performed.
<ul style="list-style-type: none"> Level-1 Limit Managed 	<ul style="list-style-type: none"> Reference Data Representativeness - Reference measurements assessed to be mostly representative of the satellite measurements, covering a primary range satellite of measurements and at ad hoc opportunities (no formal documented regular timescale). Reference Data Quality - Reference data comes a single uncertainty for the entire dataset. Validation Method - Methodology assess satellite measurements, simple uncertainty estimated e.g. from statistical spread for results. Validation Results - Validation results show good agreement between satellite and reference measurements within uncertainties in most cases.
<ul style="list-style-type: none"> Level-2 Managed 	<ul style="list-style-type: none"> Reference Data Representativeness - Reference measurements assessed to be well representative of the satellite measurements, covering a reasonable range of the satellite's measurements and carried out using FRM or community approved methods. Carried out on a regular timescale of approximately annual basis but not necessarily based on need. Reference Data Quality - Reference data comes with full uncertainty information, assessed following the GUM and traceable to community reference or SI (e.g. FRM) Validation Method - Methodology assesses satellite measurements and reference data w.r.t. their uncertainties Validation Results - Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Analysis performed independently of satellite mission owner.
<ul style="list-style-type: none"> Level-3 Well Managed 	<ul style="list-style-type: none"> Reference Data Representativeness - Reference measurements independently assessed to be fully representative of the satellite measurements, covering the satellite's full range of measurements and with full assessment of uncertainties and carried out on a regular basis determined by product performance. Reference Data Quality - Reference data comes with full uncertainty and error-correlation information, assessed following the GUM and traceable to SI (e.g. FRM). Validation Method - Methodology assess satellite measurements and reference data w.r.t. their error- covariance and validates those uncertainties. Validation Results - Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Uncertainty validated. Analysis performed independently of satellite mission owner.



Level of maturity could be defined by filling a form (i.e. responding to questions for each parameter)



	DMP-6
<ul style="list-style-type: none"> Level-0 Not Managed 	abcd = 0
<ul style="list-style-type: none"> Level-1 Limit Managed 	$1 \leq abcd \leq 40$
<ul style="list-style-type: none"> Level-2 Managed 	$41 \leq abcd \leq 80$
<ul style="list-style-type: none"> Level-3 Well Managed 	abcd = 81