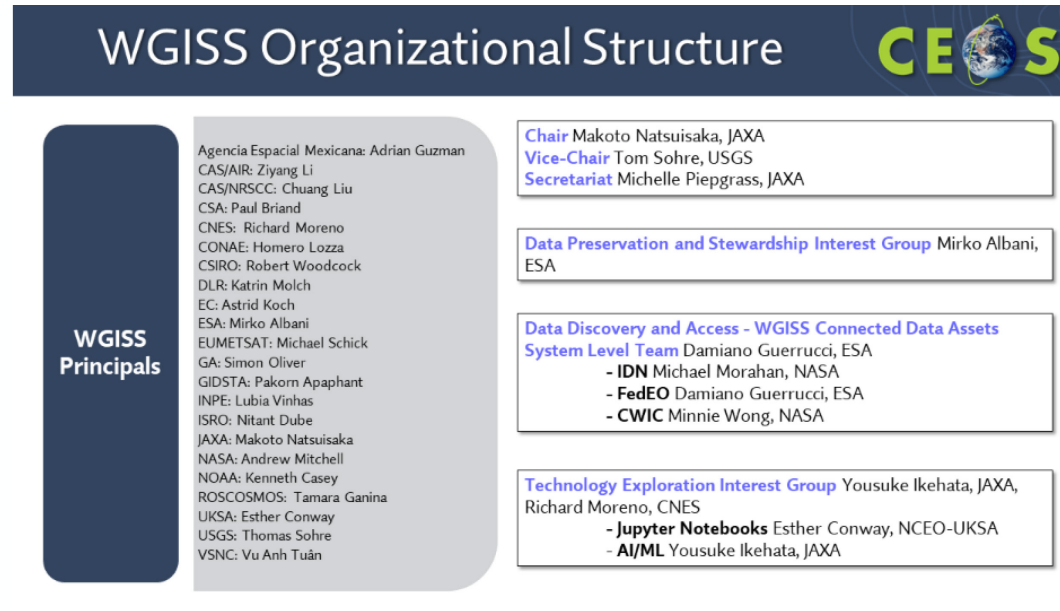


CEOS Data Management and Stewardship Maturity Matrix in support to Earth Observation data preservation and curation

I. Maggio (Rhea for ESA) OLOS September 25, 2023

WGISS (The Working Group on Information Systems and Services) is a subsidiary body supporting CEOS.



- ✓ Promotes collaboration in the development of systems and services that manage and supply Earth Observation data;
- ✓ Creates and demonstrates prototypes supporting CEOS and Group on Earth Observation (GEO) requirements;
- ✓ Addresses the internal management of EO data, the creation of information systems and the delivery of interoperable services.

The activities and expertise of WGISS span the full range of the information life cycle from the requirements and metadata definition for the initial ingestion of satellite data into archives through to the incorporation of derived information into end-user applications.

<https://ceos.org/ourwork/workinggroups/wgiss/preservation/>

The screenshot shows the CEOS website page for 'Preservation and Stewardship'. The page includes a navigation menu on the left with categories like 'Our Groups', 'Working Groups', and 'Virtual Constellations'. The main content area is titled 'Preservation and Stewardship' and contains sections for 'Background', 'Purpose', 'Scope', 'Results', and 'Contact Us'. The 'Background' section states that Earth Observation data are unique snapshots of the condition of the Earth at a specific point in time. The 'Purpose' section lists activities such as enabling the sharing of agency investigations, developing best practices, and maintaining a 'Data Purge Alert' service. The 'Scope' section focuses on EO Data, Metadata, and Associated Knowledge. The 'Results' section lists various white papers and reports, including 'Long-Term Archive Strategies' and 'Data Preservation Techniques'. The 'Contact Us' section provides information on how to reach the Data Stewardship Interest Group point of contact, Mirko Albani.

- Enable the sharing of agency investigations, developments, experiences and lessons learned relating to EO data stewardship.
- Draft common cross-agency best practices or guidelines of data stewardship for possible adoption by WGISS.
- Sponsor technical exchanges and the conduction of Joint Activities and/or Pilot Projects on specific data stewardship topics.
- Establish and maintain a CEOS “Data Purge Alert” service.
- Contribute to GEO and Standardization activities.
- Activities focus on EO Data, Metadata, and Associated Information.
- Long-term archiving approaches, systems and media.
- Data Formats and Standards.
- Preservation Lifecycle concepts.
- Data Valorization and Curation.

Three questions come to light...

What is the Maturity Matrix/Model?



Who could use it?

Why it should be used?

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

Maturity models/matrices are used to measure “levels of maturity” addressing the needs of specific domains. Examples:

- Capability Maturity Model Integration (CMMI)
- Levels of Maturity of Digital Repositories (e.g. ISO 16363)
- Climate Data Record Maturity Matrix (CDRMM)
- ESA TECHNOLOGY READINESS LEVELS (TRLs)
- ESA Scientific Readiness Levels (SRL)

Maturity Matrix for Long-Term Scientific Data Stewardship (2015, Ge Peng and Jeffrey L. Privette) covers the full scientific data lifecycle

Data providers

- to evaluate and improve the quality and usability of their products



Modelers, decision-makers, and scientists

- to improve their products
- to make investment and use decision

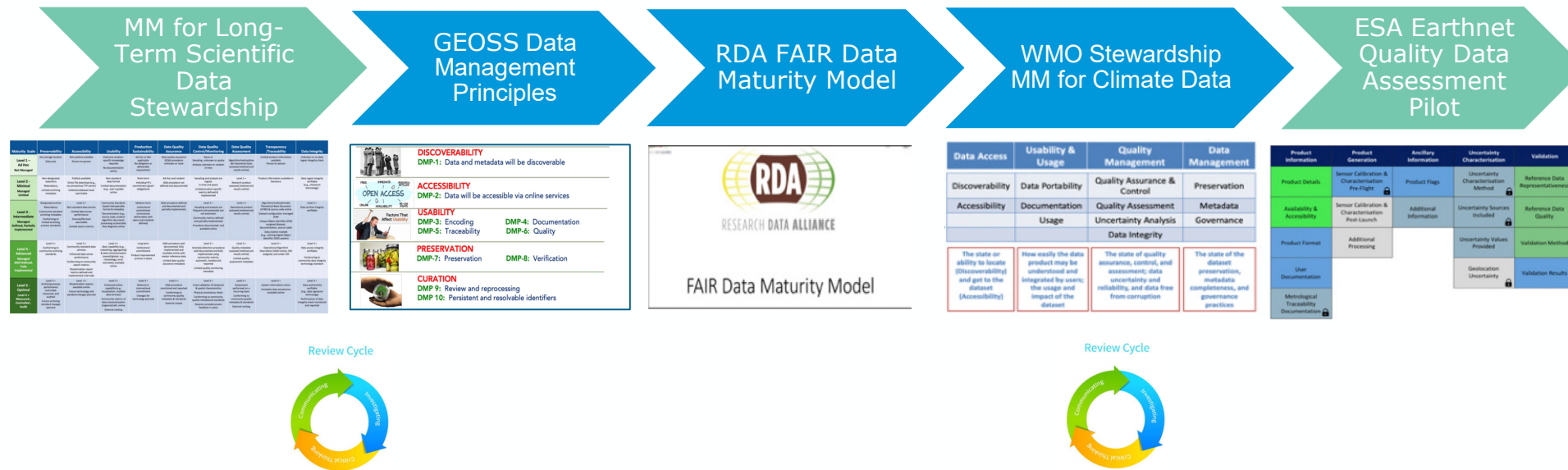
Data managers/stewards of data centers and repositories

- to validate their compliance or lack of stewardship practice or standards
- to assess the current state
- to create a roadmap forward to improve or enhance its stewardship maturity of practices applied to all its holdings

- ✓ Provides data quality, usability information to users, stakeholders, and decision makers;
- ✓ A reference model for stewardship planning and resource allocation;
- ✓ Creates a roadmap for scientific data stewardship improvement;
- ✓ Provides detailed guideline and recommendations for preservation;
- ✓ Evaluates if the preservation follows best practices;
- ✓ Gives a technical evaluation of the level of preservation and helps with self assessment of preservation;
- ✓ Gives no numbers or average but a status;
- ✓ Helps to break the problem down, and understand the costs associated with each;
- ✓ Funding agencies can define goal levels;
- ✓ Flexible and adaptable after a tailoring.

.... This is enough, isn't it???

CEOS WGISS DMSMM: Generation process



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).

5 Areas

- Discoverability
- Accessibility
- Usability
- Preservation
- Curation

4 Level of Maturity

- L0 Not Managed
- L1 Partially Managed
- L2 Managed
- L3 Fully Managed

12 Components

- Metadata for Discovery
- Online Access
- Data encoding
- Data Documentation
- Data Traceability
- Data Validation
- Data Metrology (e.g. Uncertainty)
- Data Quality Control
- Product Details
- Data Preservation
- Data Verification
- Data Processing/Reprocessing
- Persistent & Resolvable Identifier

- The goal of the DMSMM is to provide a holistic, consistent, quantifiable, and scalable measure of data stewardship maturity for end users and stakeholders including data providers and decision-makers.
- The DMSMM should be tailored for each organization and eventually specific dataset to properly take into account individual Data Stewardship and Data Management needs.
- The DMSMM helps data stewards and curators to get a consistent and quantifiable measure of an organisation's data holdings maturity.
- The ratings of DMSMM will help to validate compliance with applicable regulations on stewarding digital environmental geospatial data. The results can be used to identify potential areas for improvement and to create a roadmap for enhancing maturity of selected datasets in the identified areas by following community-accepted best practices.
- The evaluation of the DSMMM of a product can be used to build a stewardship cost model for planning purposes – based on the difference between the current maturity levels of key components and relevant stewardship requirements – prior to beginning the archive and data governance process.
- The DMSMM can be utilised by data providers or scientific stewards seeking to evaluate and improve the quality and usability of their products. The results can be also used by scientists to better understand the upstream data and data quality management practices applied to their input datasets.

Step by step in DMSMM

PRESERVATION		PRESERVATION		
	MMP9 Data Preservation	MMP10 Data Verification	MMP9 Data Preservation	MMP10 Data Verification
L0	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	integrity, authenticity and readability check	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
L1	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) 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
L2	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) 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
L3	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) 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

Perform a verification for each task putting in green the tasks already implemented.

PRESERVATION		PRESERVATION	
MMP9 Data Preservation	MMP10 Data Verification	MMP9 Data Preservation	MMP10 Data Verification
L0 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	integrity, authenticity and readability check	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
L1 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) 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
L2 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) 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
L3 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) 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

Perform this kind of verification for each task in the maturity level under analysis.

		PRESERVATION	
		MMP9 Data Preservation	MMP10 Data Verification
L0		<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> integrity, authenticity and readability check
L1		<ul style="list-style-type: none"> 1) Basic archiving for original data records preservation 2) Assessment of SW preservation 3) Product Details Assessment: Any required information missing 	<ul style="list-style-type: none"> Data Records/Associated Information integrity basic check
L2		<ul style="list-style-type: none"> 1) Preservation repository certified internally 2) Community-standard for archiving metadata 3) Product Details Assessment: All required information available, any recommended information missing 	<ul style="list-style-type: none"> 1) Data Records/Associated Information content integrity check and verification 2) Media readability and accessibility testing
L3		<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> 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

		PRESERVATION	
		MMP9 Data Preservation	MMP10 Data Verification
		<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> No Data/Associated Information integrity, authenticity and readability check
		<ul style="list-style-type: none"> 1) Basic archiving for original data records preservation 2) Assessment of SW preservation 3) Product Details Assessment: Any required information missing 	<ul style="list-style-type: none"> Data Records/Associated Information integrity basic check
		<ul style="list-style-type: none"> 1) Preservation repository certified internally 2) Community-standard for archiving metadata 3) Product Details Assessment: All required information available, any recommended information missing 	<ul style="list-style-type: none"> 1) Data Records/Associated Information content integrity check and verification 2) Media readability and accessibility testing
		<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> 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

When all tasks in the analysed maturity level are fully covered, the whole cell becomes green.



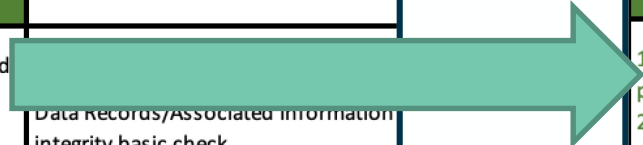
L0

L1

L2

L3

PRESERVATION	
MMP9 Data Preservation	MMP10 Data Verification
<ul style="list-style-type: none"> 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
<ul style="list-style-type: none"> 1) Basic archiving for original data record preservation 2) Assessment of SW preservation 3) Product Details Assessment: Any required information missing 	Data Records/Associated Information integrity basic check
<ul style="list-style-type: none"> 1) Preservation repository certified internally 2) Community-standard for archiving metadata 3) Product Details Assessment: All required information available, any recommended information missing 	<ul style="list-style-type: none"> 1) Data Records/Associated Information content integrity check and verification 2) Media readability and accessibility testing
<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> 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



PRESERVATION	
MMP9 Data Preservation	MMP10 Data Verification
<ul style="list-style-type: none"> 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
<ul style="list-style-type: none"> 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
<ul style="list-style-type: none"> 1) Preservation repository certified internally 2) Community-standard for archiving metadata 3) Product Details Assessment: All required information available, any recommended information missing 	<ul style="list-style-type: none"> 1) Data Records/Associated Information content integrity check and verification 2) Media readability and accessibility testing
<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> 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

The analysis continues going on to the next level of maturity.



L0

L1

L2

L3

PRESERVATION	
MMP9 Data Preservation	MMP10 Data Verification
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) 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) 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) 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



PRESERVATION	
MMP9 Data Preservation	MMP10 Data Verification
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) 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) 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) 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

When a task is not already implemented it remains in black and at the end the cell, partially covered, becomes light green.



