

# heritage data programme

## → DATA MANAGEMENT AND STEWARDSHIP MATURITY MATRIX SUPPORTING DATA CURATOR

The value of Earth observations are maximized through data life-cycle management based on ten principles supporting five themes.

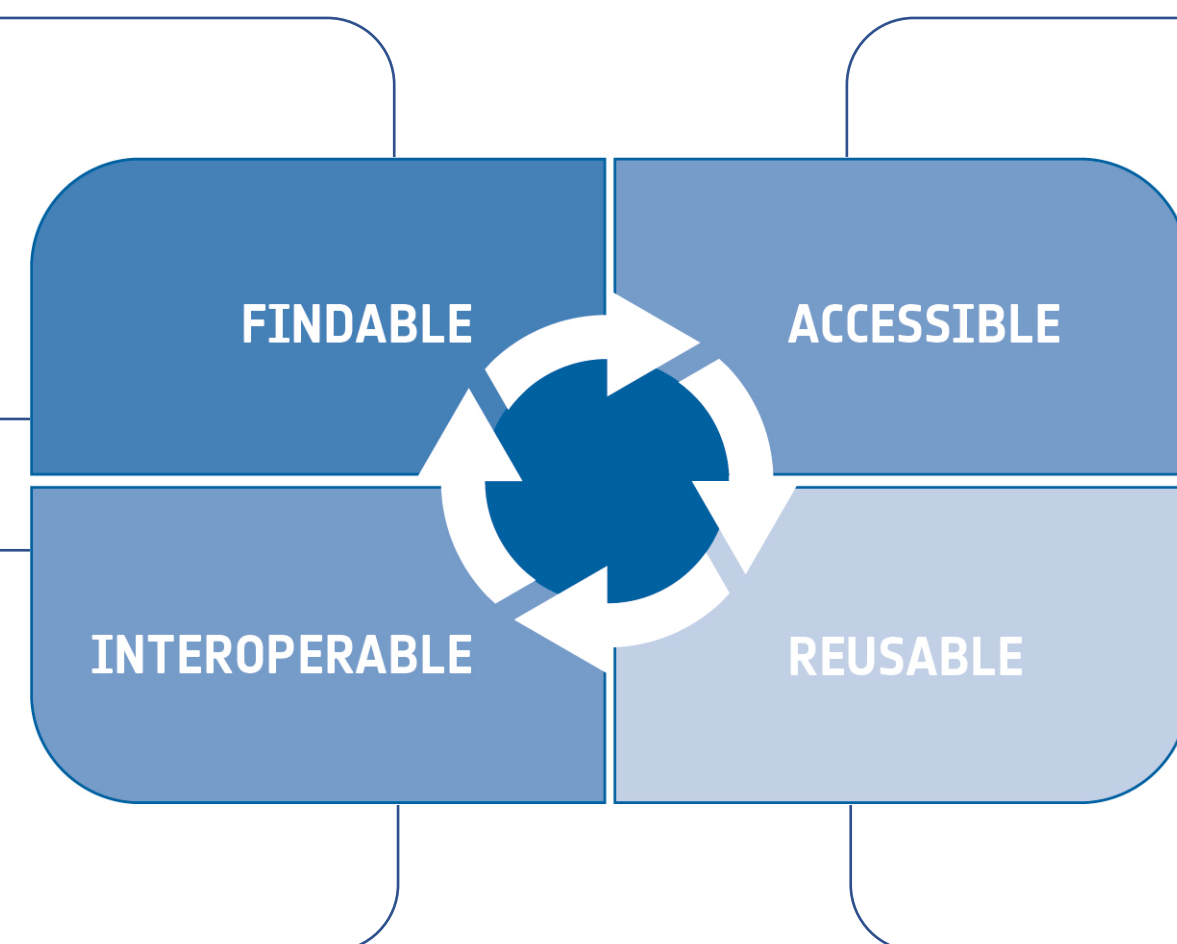
<b>DISCOVERABILITY</b> DMP-1: Data and metadata will be discoverable
<b>ACCESSIBILITY</b> DMP-2: Data will be accessible via online services
<b>USABILITY</b> DMP-3: Encoding DMP-5: Traceability DMP-4: Documentation DMP-6: Quality
<b>PRESERVATION</b> DMP-7: Preservation DMP-8: Verification
<b>CURATION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

GEO Data Management Principles Implementation Guidelines

Maturity Scale	Level 1 - Ad Hoc	Level 2 - Minimal	Level 3 - Intermediate	Level 4 - Advanced	Level 5 - Mature
Key Component	Not Managed	Managed Limited	Managed Defined, Partially Implemented	Managed Well-Defined, Fully Implemented	Level 4 + Measured, Controlled, Audit
Preservability	The state of dataset being preservable				
Accessibility	The state of dataset being publicly searchable and accessible				
Usability	The state of data product being easy to understand and use				
Production Sustainability	The state of data production being sustainable and extendable				
Data Quality Assurance	The state of data product quality being assured/screened				
Data Quality Control /Monitoring	The state of data product quality being controlled and monitored				
Data Quality Assessment	The state of data product quality being assessed				
Transparency /Traceability	The state of data product being transparent, trackable, and traceable				
Data Integrity	The state of data integrity being verifiable				

A Unified Framework for Measuring Stewardship Practices Applied to Digital Environmental Datasets

- globally unique and persistent identifier assignment and specification
- rich metadata
- registered or indexed in a searchable resource



- identifier using a standardized communications protocol
- protocol open, free and universally implementable
- authentication and authorization procedure

- formal, accessible, shared and broadly applicable language for knowledge representation
- vocabularies that follow the Fair Principles
- qualified references to other metadata

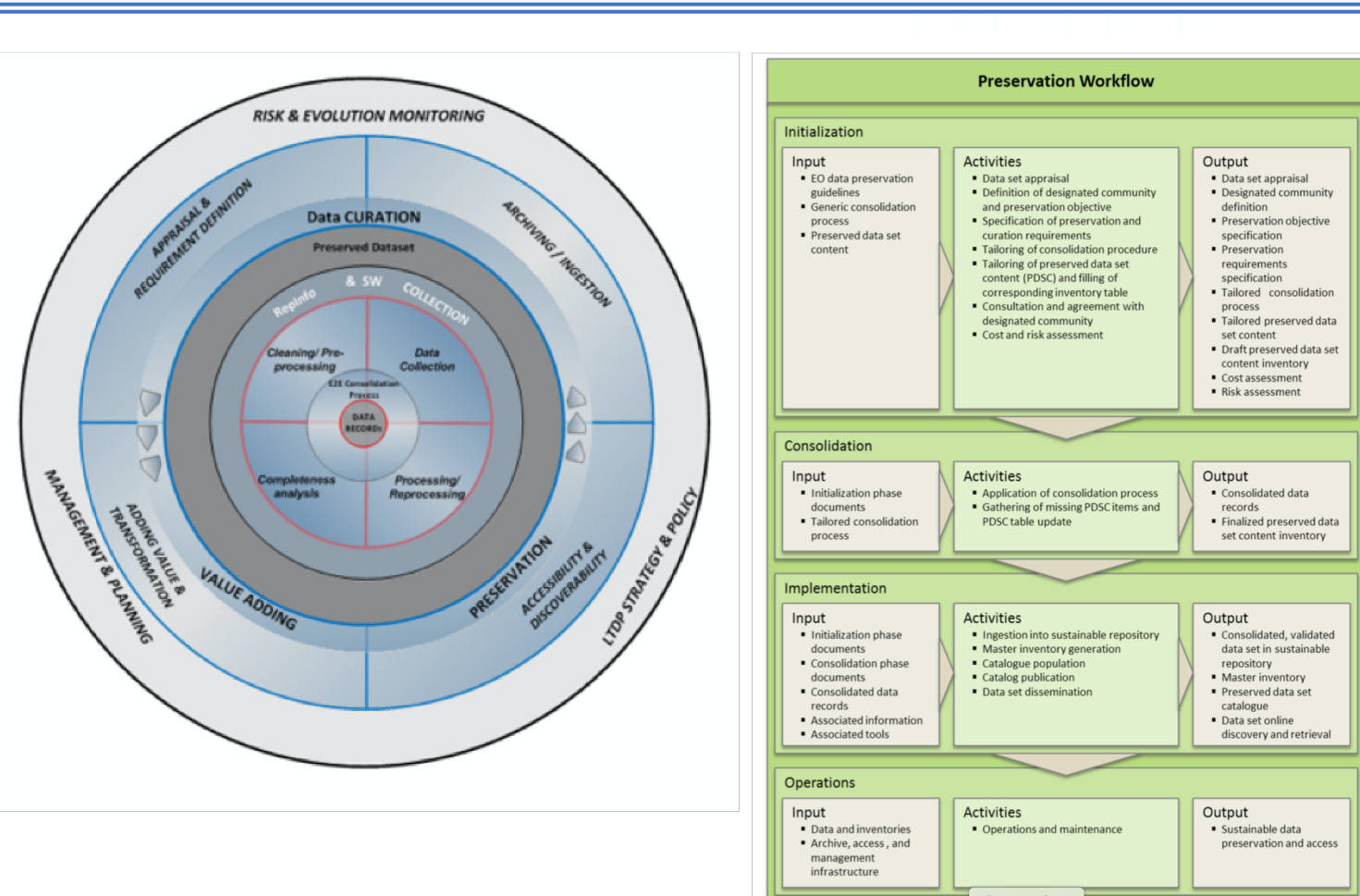
- plurality of accurate and relevant attributes
- clear and accessible data usage license
- associated with its provenance
- meet domain-relevant community standards

## DATA MANAGEMENT AND STEWARDSHIP MATURITY MATRIX

	Discoverability DMP-1 Metadata for Discovery	Accessibility DMP-2 Online Access	Usability DMP-3 Data Encoding DMP-4 Data Documentation DMP-5 Traceability DMP-6 Quality	Preservation DMP-7 Preservation DMP-8 Verification DMP-9 Reprocessing DMP-10 Persistent Identifier
<b>Level-0 Not Managed</b>	1) No catalogue available 2) No advertising available	Data are not accessible online	Data Not Structured 1) Already existing mission documentation (mission, data and product documentation)	1) No control and monitoring check 2) No quality indicator in metadata 3) No procedures documentation
<b>Level-1 Limit Managed</b>	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)	1) Basic schema for automated data use	1) Basic archiving for original data records preservation 2) The entity in charge of data long term preservation is identified and designated 3) Minimal redundancy and metadata preservation 4) Assessment of SW preservation
<b>Level-2 Managed</b>	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) 3) Link between mission documentation and data record created and managed (internal use only)	1) Documentation produced, published and well described (covering the format, metadata and methods used in creating and validating the data) 2) Quality control procedures documented and available online 3) Quality indicator post processing available 4) Quality control procedures documented and available online	1) Preservation repository certified internally 2) Documented storage procedures (planning of periodic media refreshment) 3) Redundancy managed (e.g. back-up, different media technology) 4) Basic archiving processes measured and controlled 5) Community standard for archiving metadata (e.g. AIP)
<b>Level-3 Well Managed</b>	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 disseminate 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) Visualization 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 adaptation to new technologies and standards evolution	1) Automatic metadata generation for provenance documentation (e.g. to support the reproducibility of science) 2) Link between mission documentation and data record published 3) Standards based metadata documentation (e.g. to support the reproducibility of science) 4) Complete and updated data provenance available online 5) Data quality control fully compliant with an international standard 6) Quality indicator pre and post processing available in the metadata 7) Quality metadata assessed	1) Preservation repository officially certified (e.g. ISO 15926, CoreTrustSeal) 2) Periodic technology refreshment 3) Identify and manage the basic preservation of relevant mission SW, ensuring that preserved data can be recovered 4) Continuity of service availability (Business Continuity, Disaster Recovery, etc.) 5) Automatic Data Records/Associated Knowledge content integrity check and verification 6) Data authenticity verifiable internally and by the final user 7) Automatic verification process, including monitoring and reporting



A Long Term Data Preservation Working Group (ASI, CNES, CSA, DLR and ESA) was formed at the end of 2007 within the EO Ground Segment Coordination Body (GSCB) to start cooperation activities at European level in the LTDP field and to raise awareness on the LTDP issue involving all European Earth Observation mission owners and archive holders. Cooperation activities related to heritage data management and stewardship have been extended starting in 2014 also to the Committee on Earth Observation Satellites (CEOS) in the frame of the Working Group on Information Systems and Services (WGISS). This cooperation has lead, among others, to the consolidation and issue of several CEOS Best Practices on data stewardship based on the results of the LTDP WG, which constitute the basis for ESA's and other EO data archive holders stewardship approach. In the Earth Observation domain, CEOS and GEO will continue to be the main vehicle for an international cooperation, and EO GSCB for cooperation at European level. As a continuation of the activities started as part of the LTDP programme, ESA will continue to support standardization bodies (e.g. OGC, CCSDS, ISO, ECSS) for the completion of the standardization activities related to the Preserved Data Set Content and for the inclusion of a Heritage Mission Phase as part of the ECSS standards.



In the modern era of big data the curation of data has become more prominent, particularly for processing and preserving high volume of data, and aims at the management of data throughout their lifecycle. The Preservation Workflow Best Practices is composed of:

- Appraisal & Requirements Definition;
- Acquisition & Ingestion/Archiving;
- Accessibility & Discoverability;
- Valorisation.

### PILLARS

Preservation

Discovery

Access

Exploitation/Value Adding



Data Manager

Data Curator

Data Scientist

### DATA CURATOR

#### NEW ROLE

Data curation includes "all the processes needed for a controlled data creation, maintenance, and management, together with the capacity to add value to data".

The main purpose of data curation is therefore to ensure that data are preserved, reliably retrieved for future research purposes or reuse.

Data curator makes data scientists more productive.

Use case	Preservation	Discovery	Access	Valorisation	Exploitation
Appraisal of Historical Space Data Collection					
Data Purge Alert service					
Consolidation of the Preserved Data Set Content					
Tailoring of the Preserved Data Set Content					
Definition of the preservation requirements for future Space Missions					
Apply and maintain updated all Space Data Preservation and Stewardship standards and best practices					
Ensure Space data records and associated information provenance and context					
Ingestion of associated information and definition of the relations among the EO Missions/Sensors Data Records collections and the relevant Associated Information					
Contribute with the monitoring and definition (if needed) of Space data formats specification					
Tailoring and monitoring the Space mission maturity matrix					
Definition of the Master PDSC (Certification)					
Facilitate and improve discoverability and accessibility of Space Mission Data and Associated Information (e.g. including DOI assignment)					
Analysis of collected metrics to better address funding and improve decision making (e.g. user community behaviour, systems and services performances, etc.)					
Verification of the Space Data Collection Quality Indicator					
Perform thematic analysis for time series, FCDR generation and reprocessing campaign					
Correlate Space Mission Data with in-situ data					