
Long Term Preservation of Earth Observation Space Data

Preservation Guidelines

CEOS-WGISS

Data Stewardship Interest Group

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

Date: September, 2015

Issue: Version 1.0

Document Status Sheet

Issue	Date	Comments	Editor
1.0	15 September 2015	New document evolved from European LTDP Common Guidelines Issue 2.0	M. Albani, K. Molch, I. Maggio, R. Cosac

Table of Contents

1. INTRODUCTION	1
1.1 Intended Audience	1
1.2 Background	1
1.3 Scope of Document	1
1.4 Acronyms	2
1.5 Reference Documents	2
1.6 Guidelines Application	4
1.7 Related Standards and Guidelines	4
2. THEME 1: PRESERVED DATA SET CONTENT DEFINITION AND APPRAISAL	5
3. THEME 2: ARCHIVE OPERATIONS AND ORGANIZATION	8
4. THEME 3: INFORMATION SECURITY	10
5. THEME 4: DATA INGESTION	11
6. THEME 5: ARCHIVE AND DATA MAINTENANCE	13
7. THEME 6: DATA ACCESS AND INTEROPERABILITY	15
8. THEME 7: DATA EXPLOITATION AND RE-PROCESSING	17
9. THEME 8: DATA PURGE PREVENTION	20
ANNEX A – GUIDELINES PRIORITY AND ASSOCIATION WITH RELATED DATA MANAGEMENT STANDARDS AND PRINCIPLES	21

List of Figures

Figure 1: Mapping of the Preservation Guidelines themes (center) vs. the GEO Data Management Principles topics	22
---	-----------

List of Tables

Table 1: Applicable Documents	2
Table 2: Reference Documents	3
Table 3: Levels of adherence to the Data Preservation Guidelines	4
Table 4: GEO Data Management Principles	22
Table 5: Guidelines priority levels and relationships	25

1. INTRODUCTION

1.1 Intended Audience

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

1.2 Background

Earth Observation data are unique snapshots of the condition of the Earth at a specific point in time. As such they constitute a humankind asset, which needs to be preserved, i.e. safeguarded against loss and kept accessible and useable for current and future generations. This task becomes more important in the view of over 40 years' worth of data available in Earth Observation archives around the world – and the increasing demand for monitoring long-term variations of environmental parameters, such as sea surface temperature or global ozone distributions, which require long time series of data. Moreover, with the advent of new, high resolution Earth Observation missions and programs, data volumes are expected to grow significantly over the next years.

However, the main challenge is not the sheer volume of data, but its diversity, e.g. in format and type. Historic EO data, in particular, may be stored in different formats on various types of – possibly out-of-date media. Recovery, reformatting and reprocessing of such data, as well as the recuperation of the associated knowledge – whether it is representation information for structural and semantic understanding, mission documentation for context, or visualization and processing capacity – it becomes problematic if attempted many years after the mission has ended.

Therefore, data stewardship is the responsibility to curate the Earth observation data during all mission stages, starting with the mission planning phases and extending beyond the mission lifetime, when the - then 'historic' - data have to be kept accessible and useable for an – ideally – unlimited timespan.

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

1.3 Scope of Document

The Preservation Guidelines document covers the planning and implementation steps of the CEOS Preservation Workflow [AD-1]. The guidelines and the underlying data preservation approach should be applied to historic, current and future Earth Observation data sets. The document addresses technical and organizational aspects for the long-term preservation of EO data. It includes security, accessibility and usability aspects, recommending applicable standards and procedures.

The Preservation guidelines are not intended to cover programmatic, regulatory or data policy aspects associated with the EO data to be preserved.

Some of the guidelines are supported by a set of technical procedures, methodologies or standards, providing technical details on the recommended practical implementation.

The document addresses eight main “themes” consisting of “guiding principles” and a set of “guidelines” that should be applied to guarantee the preservation, accessibility, and usability of EO space data in the long term. The eight themes are the following:

1. Preserved data set content definition and appraisal
2. Archive operations and organization
3. Archive security
4. Data ingestion
5. Archive maintenance
6. Data access and interoperability
7. Data exploitation and re-processing
8. Data purge prevention

1.4 Acronyms

Definitions of acronyms and terms are provided in the document 'Definitions of Acronyms and Terms' [RD-2].

1.5 Applicable and Reference Documents

ID	Resource
[AD-1]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data – Preservation Workflow – CEOS/WGISS/DSIG/PW
[AD-2]	Long Term Preservation of Earth Observation Space Data - Earth Observation Preserved Data Set Content, CEOS/WGISS/DSIG/EOPDSC
[AD-3]	CEOS Best Practices on Long Term Preservation of Earth Observation Space Data - EO Data Stewardship Definition
[AD-4]	CEOS EO Data Purge Alert Procedure, http://wgiss.ceos.org/purgealert/

Table 1: Applicable Documents

ID	Resource
[RD-1]	European LTDP Common Guidelines 2.0, GSCB-LTDP-EOPG-GD-09-0002, June 2012
[RD-2]	Definitions of Acronyms and Terms v0.7, January 2015, CEOS/WGISS/DSIG/DEF
[RD-3]	Towards Data Management Principles, Data Management Principles Task Force, GEO, November 2014, https://www.earthobservations.org/documents.php?smid=100
[RD-4]	Quality Assurance Framework for Earth Observation - Guidelines Framework (QA4EO) www.qa4eo.org
[RD-5]	Producer Archive Interface Specification (PAIS): CCSDS 651.1-R-1 http://public.ccsds.org/review/Poll-jgg-20111223
[RD-6]	ISO 14721:2012 - Space data and information transfer systems - Open archival information system (OAIS). CCSDS 650.0-M-2
[RD-7]	ISO 27001:2013 - Information security management systems - Requirements
[RD-8]	ISO 20652:2006 - Space data and information transfer systems - Producer Archive interface - Methodology Abstract Standard (PAIMAS). CCSDS 651.0-M-1
[RD-9]	ISO 16363:2012 - Space data and information transfer systems - Audit and certification of trustworthy digital repositories. CCSDS 652.0-M-1
[RD-10]	Heterogeneous Mission Accessibility (HMA) specifications and Cookbook Version 0.9.9.7, 2 February 2012
[RD-11]	CEOS OpenSearch Best Practice Document, Version 1.0.1, CEOS-OPENSEARCH-BP-V1.0.1

Table 2: Reference Documents

1.6 Guidelines Application

The guidelines have been classified into different levels of priority. The application of the guidelines could follow a step-wise approach starting with implementing the high priority ones. This will yield Level A adherence, which guarantees a basic level of security, integrity and accessibility of the archived data (Table 3).

Adherence Level	Description	Condition for adherence
Level A	Basic data security, integrity and access.	Implementation of all high priority guidelines.
Level B	Medium-level data security, integrity, access and interoperability.	Implementation of all high and medium priority guidelines.
Level C	Top-level data security, integrity, access, and interoperability.	Implementation of all guidelines.

Table 3: Levels of adherence to the Data Preservation Guidelines

A self-assessment should be done in order to determine the level of adherence to the Preservation Guidelines. A subsequent impact analysis assists in determining the effort and cost of increasing the data holder's level of adherence to the Guidelines. A dedicated table is available to assist the data holders in performing both the self-assessment and the impact analysis.

1.7 Related Standards and Guidelines

Where applicable in the Earth Observation context, the Preservation Guidelines refer to additional standards or consolidated approaches (“de facto standards”). Furthermore, the ISO-standard on 'Space data and information transfer systems - Audit and certification of trustworthy digital repositories' (ISO 16363:2012) and the 'Data Management Principles' of the Group on Earth Observations (GEO) [RD-3] also provide guidelines useful for the preservation and management of Earth Observation data.

The relationship between the Preservation Guidelines and ISO 16363 has been analysed. The analysis has shown very good compatibility between the two approaches, in particular concerning the technical aspects. ISO 16363 is an audit and certification method which addresses the trustworthiness of all types of digital repositories. The Preservation Guidelines on the other hand are a set of practical recommendations specifically targeted at Earth Observation data holdings and covering largely the topics addressed in the standard. Since the Data Preservation Guidelines are not addressing on purpose financial, policy and legal/organizational aspects of Earth Observation archives, a full matching with ISO 16363 cannot be achieved. However, adherence to the Preservation Guidelines can be considered as a first valuable step towards the certification of an Earth observation digital repository against ISO 16363.

The Group on Earth Observations (GEO) has drafted a set of high level common Earth observation Data Management Principles covering the entire data life cycle from planning, to acquisition, quality assurance, documentation, access, archiving, and preservation. This will ensure that data and information of different origin and type are comparable and compatible, facilitating their integration into models and the development of applications to derive decision support tools. A comparison has been done between the Preservation Guidelines and the Data Management Principles of the Group on Earth Observations (GEO).

The results of both comparative analyses are provided in Annex 0

2. THEME 1: PRESERVED DATA SET CONTENT DEFINITION AND APPRAISAL

Rationale

A structured appraisal of a data set considered for long-term preservation helps assess if an Earth Observation data set is valuable enough to be preserved for future generations. It also provides an estimate of the cost and effort involved, and of any associated risks.

In order to ensure that a preserved data set remains fully understandable and useable in the future, additional information - beyond the instrument data and the metadata - needs to be preserved as well. This 'associated knowledge' can include e.g. information on the structure and semantics of the data set, on calibration, or processors. The 'Preserved Data Set Content' provides a mission-specific list of information and tools to be preserved along with the instrument data to ensure long-term usability.

Description

The “Preserved Data Set Content” represents the complete set of data and information to be preserved, required for enabling current and future understanding and utilization of the preserved data by a designated user community. The goal is to preserve and make accessible this complete dataset including information required to determine “Quality Indicators” as required by [RD-4].

Guidelines

GUIDELINE 1.1 – Preserved Data Set Content - (Level A)

Identify and archive the following set of data content for each Earth Observation space mission or instrument¹ according to the “Earth Observation Preserved Data Set Content” document [AD-2]:

- a) *Data Records*: these include Raw data², Level 0 data and higher-level products³, browses, auxiliary and ancillary data, calibration and validation data sets⁴, and metadata.
- b) *Processing Software*: this includes all the software used in the product generation, quality control, and the product visualization and value adding tools.
- c) *Mission Documentation*⁵: this includes among others mission architecture, products specifications, instruments characteristics, algorithms description, Cal/Val procedures, mission/instruments performances reports, quality related information [RD-4], etc.

The “Earth Observation Preserved Data Set Content” document [AD-2] can be tailored to meet specific requirements of the mission data set to be preserved.

- ✓ Earth Observation Preserved Data Set Content [AD-2]
- ✓ Quality Assurance Framework for Earth Observation (QA4EO) [RD-4]

¹ Identification and archiving of all this information will probably not be achievable for old missions. A tailoring between what is required and what is available or generated for a mission (past, current and future) will be in any case necessary.

² Raw data shall be preserved whenever conversion to Level 0 cannot be adequately certified.

³ When systematically generated as part of the mission requirements and/or reprocessed.

⁴ Including processing/reference validation data sets.

⁵ Mission Documentation shall include Representation Information, Packaging Information and Preservation Descriptive Information according to OAIS information model [RD-6].

GUIDELINE 1.2 – Preserved data set inventory - (Level A)

Generate and maintain a complete inventory of the archived Preserved Data Set Content defined for each mission/instrument with the following items as a minimum:

- ✓ Description and availability of the data records, processing software and mission documentation (specifying all elements identified in guideline 1.1).
- ✓ Time span and volumes of the data records.
- ✓ Physical locations of the data records.
- ✓ Media of storage and archive formats.
- ✓ Processing Software (if maintained or simply archived) information: versioning, IPR/licenses, etc.
- ✓ Mission Related Documentation: versioning, repository, etc.
- ✓ Links between Data Records and relevant Quality Information.

GUIDELINE 1.3 - Bi-directional linkages - (Level B)

Create bi-directional logic linkages between archived Data Records and the associated Mission Documentation (e.g. Sensor Specification document, Ground Segment Specification document, Quality Information etc.) and Processing Software necessary to understand and use it.

GUIDELINE 1.4 - Preserved data elements - (Level B)

Assess and harmonize the format of all the “preserved data set content” elements (see Guideline 1.1) with the archive's standards in order to make the adopted formats understandable and sustainable.

- | |
|--|
| ✓ PAIS (Producer Archive Interface Specification): implementing PAIMAS standard (in the process of becoming a CCSDS Blue Book), [RD-5] |
|--|

GUIDELINE 1.5 – Archived data records format - (Level B)

Adopt a common standard archive format (AIP) for data records.

- | |
|---|
| List of recommended archive formats:
✓ SAFE – Standard Archive Format for Europe (http://earth.esa.int/SAFE/index.html) |
|---|

GUIDELINE 1.6 – Archived data records exchange format – (Level B)

Adopt a common standard format for the exchange of archived data.

GUIDELINE 1.7 – Mission Documentation format - (Level B)

Adopt a common Mission Documentation standard format suitable for long-term preservation.

- | |
|---|
| List of recommended formats:
✓ Portable Document Format/ Archiving (PDF/A)
✓ Flexible Image Transport System (FITS) |
|---|

GUIDELINE 1.8 – Data appraisal procedure - (Level A)

Perform a “Data Appraisal” procedure to assess and document the value and prospects of each Preserved Data Record.

3. THEME 2: ARCHIVE OPERATIONS AND ORGANIZATION

Rationale

A professionally organized, largely automated archive infrastructure, operated by skilled personnel, ensures that the preserved Earth Observation data remain technically safe, secure and well maintained.

Description

Archive operations consist of all daily activities which are carried out to run and monitor the archive system (execution and control of the applications, system monitoring, anomaly reporting, error recovery, activity reporting and statistics, etc.). An appropriate organizational structure ensures that the tasks and processes of a digital long-term archive are performed in a professional and sustainable manner.

Guidelines

GUIDELINE 2.1 – Reference model for archive - (Level A)

Adopt a common standard reference model for the archives.

- | |
|---|
| ✓ ISO 14721 - OAIS standard (ISO reference model for Open Archival Information System) [RD-6] |
|---|

GUIDELINE 2.2 – Operations procedures - (Level A)

Perform archive operations following a set of approved and consolidated documented operational procedures.

GUIDELINE 2.3 – Archives equipment maintenance - (Level A)

Keep archives equipment in conformance with manufacturer recommendations.

GUIDELINE 2.4 – Archives automation - (Level A)

Implement automation of the archives (e.g. utilizing automatic robot libraries and software to ensure homogeneous tasks performances) to minimize the number of operations requiring operators' intervention.

Guideline 2.5 – Archives organisation - (Level A)

Establish an appropriate archive organisational structure based on a sufficient number of qualified staff with clear roles and responsibilities. Archive operation is governed through the organisational structure that oversees the planning and operation of preservation tasks.

Guideline 2.6 – Archives legal and contractual aspects - (Level B)

Legal aspects and contractual rules related to data ingestion, archive operations and data access and dissemination should be in place.

Guideline 2.7 – Archive configuration management - (Level A)

Maintain under configuration control the archived Data Records, Mission Documentation and Processing Software including their internal dependencies and linkages.

GUIDELINE 2.8 – Archive AIPs identifiers - (Level A)

Define and use a convention that generates unique identifiers for Archive Information Packages (AIPs).

GUIDELINE 2.9 – Archive AIPs persistent identifiers - (Level C)

Define and use a convention that generates unique **persistent** identifiers for Archive Information Packages (AIPs).

GUIDELINE 2.10 - Archive preservation planning and actions - (Level A)

Generate and maintain one or more documents describing the adopted preservation strategy, planning and actions relevant to the archive holdings, and clearly defining current compliance to the Data Preservation guidelines and future plans to improve adherence.

GUIDELINE 2.11 – Archive content monitoring - (Level C)

Put in place and maintain mechanisms for monitoring the understandability and usability of the archive content.

GUIDELINE 2.12 – System infrastructure risks - (Level A)

Identify and manage the risks associated with system infrastructure which could affect the preservation objectives (e.g. monitor end of life of technologies).

GUIDELINE 2.13 – Archiving systems common approach - (Level C)

Pursue a harmonized approach within the Earth Observation community for the future development of archiving systems to improve compatibility of services provided by different organizations (e.g. exchange of specifications, application software, best practices, etc.).

4. THEME 3: INFORMATION SECURITY

Rationale

Implementing a comprehensive information security framework, based e.g. on the ISO 27000 family of standards, ensures physical safety, integrity and the appropriate confidentiality of the preserved data sets.

Description

This theme encompasses all the activities dedicated to the implementation of security measures for storage and access to the content of an archive.

Guidelines

GUIDELINE 3.1 – Archives security requirements - (Level B)

Base archive security requirements on international standards and policies. The archive infrastructure should put into practice the specifications for handling the preserved data and content on the technology and security levels.

✓ ISO 27001:2013 - Information security management - Requirements [RD-7]
--

GUIDELINE 3.2 – Controlled access to archive facilities - (Level A)

Implement controlled access to facilities, sites and equipment to avoid physical intrusion by unauthorised persons. Allow access to core functions only to identified personnel provided with appropriate security clearances.

GUIDELINE 3.3 – Local risk mitigation infrastructure - (Level A)

Implement local risk mitigation infrastructure measures to safeguard the archives from external factors (e.g. floods, fire, disasters in general).

GUIDELINE 3.4 – Protection from external intrusion - (Level A)

Implement security mechanisms to avoid external intrusion that could harm core equipment functionalities and cause information loss.

GUIDELINE 3.5 – Information loss risk mitigation infrastructure - (Level A)

Implement measures to protect core equipment functionalities and mitigate against the risk of information loss as a consequence of internal unintentional or deliberate human actions, and of technical imperfection.

5. THEME 4: DATA INGESTION

Rationale

A harmonized, well documented approach for ingesting data sets into the archive for preservation and for generating the Archival Information Packages (AIPs) from the Submission Information Packages (SIPs) - based on international standards and recommendations - ensures the coherency and compatibility of the data sets stored in the archive and facilitates data exchange.

Description

Data Ingestion encompasses the services and functions that, according to OAIS standard [RD-6] accept Submission Information Packages (SIPs) from data producers, prepare Archival Information Packages (AIPs) for storage, and ensure that the Archival Information Packages and their supporting descriptive information are stored in the archive system.

Guidelines

GUIDELINE 4.1 – Data ingestion process - (Level A)

Carry out data ingestion according to relevant standards with documented tailoring and definition derived from the generic activities described in the standards.

- ✓ ISO 20652 - PAIMAS Standard (Producer Archive Interface Methodology Abstract Standard), [RD-8]
- ✓ PAIS (Producer Archive Interface Specification): implementing PAIMAS standard (in the process of becoming a CCSDS Red Book) [RD-5]

GUIDELINE 4.2 – Metadata generation - (Level A)

Define a descriptive set of metadata for the archived data, and generate them and relationships between the data items as part of the ingest process, or when archived data content is updated. The resulting metadata should be formatted according to relevant standards.

- List of recommended standards:
- ✓ EO collection metadata: ISO 19115 Geographic Information - Metadata [RD-10]
 - ✓ EO product metadata: OGC's GML 3.2.1 Application Schema for EO Products (OGC-07-036) [RD-10]
 - ✓ OAIS Information Model for Descriptive Information (e.g. Quality Information) [RD-6]

GUIDELINE 4.3 – Routine quality check - (Level A)

Perform routine quality check on acquired data before ingestion in the archive.

- ✓ ISO 20652 - PAIMAS Standard (Producer Archive Interface Methodology Abstract Standard) [RD-8]
- ✓ PAIS (Producer Archive Interface Specification): implementing PAIMAS standard (in the process of becoming a CCSDS Red Book) [RD-5]

GUIDELINE 4.4 – Transcription quality check - (Level A)

Perform quality check on acquired data after transcription on media in the archive to verify correct transcription of the acquired data on the media.

GUIDELINE 4.5 – Ingestion process record - (Level B)

Document the ingestion process of all archive content (as defined in guideline 1.1). Examples include: number of ingested products, number of products discarded, total size, archiving path, errors, etc.

6. THEME 5: ARCHIVE AND DATA MAINTENANCE

Rationale

Constant and careful maintenance or curation of the preserved data sets and of the software and hardware environment is necessary to ensure the integrity - and thus the usability - of the archived data over the long term.

Description

Data and archive maintenance or curation consists of all the activities aimed at guaranteeing the integrity of the archived data. Data integrity assures that the archived data are complete and unaltered through loss, tampering or data corruption. Archive maintenance refers to the storage of equipment, media and hard disk arrays in secured and environment controlled rooms, and a set of defined activities to be performed on a routine basis, such as migration to new systems and media, in accordance to the technology and consumer market evolution, data compacting and data format/packaging conversion.

Guidelines

GUIDELINE 5.1 – Archived data refreshment - (Level A)

Perform periodically a migration of the archived data (“media refreshment”) to the most adequate proven technology⁶ for data storage ensuring data access preservation⁷.

GUIDELINE 5.2 – Archived data formats description - (Level B)

Provide formal description of old archiving formats to allow the conversion to new standard formats, which will increase technical compatibility and reduce diversity of formats and interfaces between archives.

GUIDELINE 5.3 – Archived data repackaging/reformatting - (Level C)

Perform archived data repackaging and/or reformatting to comply with new standard formats and/or exchange formats to increase technical compatibility and to reduce diversity of formats and interfaces between archives⁸.

GUIDELINE 5.4 – Archived data duplication

Maintain identical copies of all archived data applying one of the security levels defined below:

- a) Dual copy in the same geographical location (but different buildings) to avoid data loss due to media degradation or obsolescence. **(Level A)**

or

⁶ Technology selection should not only be based on technical and cost aspects but also aim at the minimization of environmental impact (e.g. in terms of power consumption, thermal dissipation, etc.).

⁷ Currently data and system migrations are performed at least every five/six years.

⁸ Repackaging and/or reformatting should be performed together with archive media migration.

- b) Dual copy in the same geographical location (but different buildings) based on different technology to avoid technology based principle failures. This guideline extends Guideline 5.4.a. **(Level B)**

or

- c) Dual copy in two different geographical locations to safeguard the archive from external hazards (e.g. floods and other natural disasters, technological hazards, etc.). This guideline extends Guideline 5.4.a. **(Level B)**

or

- d) Dual copy in two different geographical locations, based on different technologies to avoid technology based principle failures. This guideline extends Guideline 5.4.b and 5.4.c **(Level C)**

GUIDELINE 5.5 – Archive system components migration (hardware) - (Level A)

Perform periodical migration of archive system components to new hardware platforms⁹.

GUIDELINE 5.6 – Media readability and accessibility tests - (Level B)

Perform periodical test for media readability and accessibility on a representative set of the archived data.

GUIDELINE 5.7 - Archive content integrity - (Level B)

Periodically verify the integrity of the archive collection/content through integrity check on a representative set of the archived data.

GUIDELINE 5.8 – Data content integrity - (Level A)

Ensure that the content of the archived data and associated information remains unchanged and, if changes are made, that these are documented and that this documentation is preserved and made available as well (provenance information).

GUIDELINE 5.9 – Obsolete media disposal - (Level A)

Organize the disposal of obsolete media in conformance with national and international environmental regulations.

⁹ Currently it is recommended to perform data and system migrations are at least each five/six years.

7. THEME 6: DATA ACCESS AND INTEROPERABILITY

Rationale

Data sets should be preserved for the sake of the user community. Thus, they should be discoverable and – within the scope of the associated data policies – be made accessible to the user community, ideally using standardized protocols to maximise interoperability and dissemination.

Description

Data access corresponds to the services and functions which make the archival information holdings and related services visible to consumers. Interoperability is related to the possibility of accessing data in a common and standardized way despite the intrinsic differences between the data sets on one hand and the accessed systems on the other hand.

Guidelines

GUIDELINE 6.1 – Preserved data set content discovery - (Level A)

Ensure continuous EO missions' preserved data set content discoverability through the following activities:

- a) Provide and maintain mechanisms to search and discover archived data records.
- b) Provide and maintain a searchable metadata and browse image catalogue of archived data records.
- c) Provide and maintain mechanisms to search and discover Mission Documentation and value adding/visualization tools relevant for the designated user community.
- d) Provide and maintain mechanism to link the EO Data Records to Quality Information.

GUIDELINE 6.2 – Preserved data set content dissemination - (Level A)

Provide and maintain preserved data set content dissemination capabilities through ordering and delivery and/or direct access (without ordering).

GUIDELINE 6.3 – On-line access and delivery - (Level B)

Ensure on-line direct access and on-line delivery services for the preserved data set content.

GUIDELINE 6.4 – Preserved data set content access and use conditions - (Level A)

Provide transparency and visibility of preserved data set content access and use conditions to users.

GUIDELINE 6.5 – Controlled preserved data set content access and dissemination - (Level A)

Implement preserved data set content controlled access and dissemination mechanisms in accordance to applicable access and use conditions (see guideline 6.4) to avoid unauthorized visibility and access.

GUIDELINE 6.6 – Preserved data set content discovery and access interfaces - (Level B)

Adhere to standard interfaces, services and delivery formats (See Guideline 6.9) to allow easy and cost-effective discovery, access and dissemination of heterogeneous EO mission (different missions, new & old) preserved data set content: interfaces for discovery, catalogue access, ordering, access and dissemination, user management and administration, etc.

- List of standards [RD-10]:
- ✓ Collection and service discovery (Advertisement): OGC’s Cataloguing of ISO Metadata using the ebRIM profile of CS-W (OGC 07-038)
 - ✓ Catalogue Service: OGC’s Catalogue Services Specification Extension Package for ebRIM Application Profile: Earth Observation Products (OGC 06-131)
 - ✓ Ordering from Catalogue: OGC’s Ordering Services for Earth Observation Products (OGC 06-141)
 - ✓ Feasibility Analysis (Programming): OGC’s Sensor Planning Service Application Profile for EO Sensors (OGC 07-018)
 - ✓ Online Data Access: OGC’s WMS EO Extension (OGC 07-063), OGC WCS extension for EO.
 - ✓ Identity (User) Management: OGC’s User Management Interfaces for Earth Observation Services (OGC 07-118)
 - ✓ CEOS OpenSearch Best Practice Document 1.0.1 [RD-11]
 - ✓ OGC 10-032r8 Geo and time Extension
 - ✓ OGC 13-026r5 Extension for Earth Observation

GUIDELINE 6.7 – Harmonization of delivery formats - (Level B)

Pursue harmonization of EO products and delivery formats (Dissemination Information Packages, DIPs) specifications (quality and content) and delivery formats (e.g. TIFF, GeoTIFF) for different missions.

GUIDELINE 6.8 - EO products realignment - (Level C)

Realign products’ (Dissemination Information Packages, DIPs) characteristics and delivery format of historical missions to established¹⁰ harmonized ones.

GUIDELINE 6.9 – Archive search capability - (Level C)

Enhance archives search capability and harmonize features extraction methods by contents, Quality Information and metadata values (e.g. Quality Thresholds).

Guideline 6.10 – Authenticity - (Level B)

Apply policies and procedures that enable the dissemination of EO products that are traceable to the source data, with evidence supporting their authenticity.

¹⁰ i.e. used by a large community of users for a sufficient period of time.

8. THEME 7: DATA EXPLOITATION AND RE-PROCESSING

Rationale

Archived data represent a unique information source in the long term and can provide valuable input to exploitation programmes. To facilitate the use of archived data and to respond to reprocessing requirements, the capability to generate missions' products should be maintained during the mission's lifetime and after the mission has ended.

Description

This theme covers activities related to the exploitation of archived data by data processing and reprocessing, regeneration or enhancement of the catalogues (e.g. through data mining), integration of new services (e.g. through service work-flow orchestration) and quality assessment of the products and services.

Guidelines

GUIDELINE 7.1 – Products generation active missions - (Level A)

Provide and maintain products generation capability (systematic or through ordering) including maintenance of the processing chains and quality control tools for active EO missions. This includes the validation of models, algorithms and software.

GUIDELINE 7.2 – Products generation non-active missions - (Level B)

For non-active missions, provide and maintain missions' products generation capability (systematic or through ordering) including maintenance of the processing chains and quality control tools, or as an alternative, perform, before dismantling the processing chains, bulk generation and archiving of all products levels according to mission requirements.

GUIDELINE 7.3 – Processing software¹¹ environment - (Level B)

Actively monitor the evolution of the environment (e.g. hardware and software operating systems) used to run "Processing Software" and perform hardware migration and software porting when necessary to maintain the original capabilities. Particular focus is on free and open source software which might guarantee higher reproducibility and robustness.

GUIDELINE 7.4 – Processing software¹¹ virtualization - (Level C)

Virtualize mission specific Processing Software to minimize dependency on the underlying environment (e.g. hardware and software operating systems).

¹¹ Including quality control, product visualization and value adding tools

GUIDELINE 7.5 – COTS software monitoring - (Level B)

Actively monitor the evolution and availability of commercial software (COTS) used as processing chains libraries and for implementation of missions' products visualization and value adding processing. Apply the countermeasures defined in the risk assessment phase.

GUIDELINE 7.6 - Reprocessing - (Level A)

Provide and maintain the capability to reprocess archived data records to generate new coherent versions of missions' products according to missions' requirements and strategy¹². Products obtained with previous algorithm versions should be maintained after reprocessing depending on missions' requirements and strategy.

GUIDELINE 7.7 – Processing/reprocessing capacity for long term data series - (Level B)

Provide the processing/reprocessing capacity to respond to missions' requirements and to projects requiring long-term data series.

GUIDELINE 7.8 – Higher level applications - (Level C)

Provide complete reference test data sets to expert users, including Quality Indicators, to facilitate the development of higher-level applications (e.g. for information extraction).

GUIDELINE 7.9 – Earth Observation data/products quality - (Level A)

Provide Quality Assurance of EO space data during the mission lifetime (e.g. through the application of international standards or guidelines) using Quality Indicators and performing validation of models, algorithms and software.

✓ A Quality Assurance Framework for Earth Observation - Guidelines Framework (QA4EO), endorsed by CEOS in November 2008, www.qa4eo.org , [RD-4]
--

GUIDELINE 7.10 – Facilitation of data exploitation - (Level C)

Pursue simplification of the workload for the users by reducing their global data handling time and cost through implementing the following measures:

- 1) Data adaptation to specific post-processing and applications.
- 2) Hosting and executing user algorithms.
- 3) Providing capability for data fusion across EO sensors.
- 4) Providing the capability to perform across disciplines data searches and integration (e.g. Earth Observation, in-situ geologic data, etc.)

¹² E.g. when new approved algorithms or auxiliary data are available or according to user community needs

GUIDELINE 7.11 – Information extraction - (Level C)

Allow information extraction from EO products through data mining and value adding services.

GUIDELINE 7.12 – Facilitate data citation - (Level B)

Assign persistent, resolvable identifiers to preserved data sets that are accessible to users, in order to facilitate data citation and encourage re-use.

9. THEME 8: DATA PURGE PREVENTION

Rationale

EO data are unique snapshots of the condition of the Earth or atmosphere and cannot be recorded again. These observations are a humankind asset fundamental for future science, targeted at better understanding our planet. Therefore, any loss of Earth Observation data should be prevented. A data holder, not able to continue preserving an Earth Observation data set, should avoid the loss of valuable data by following the steps recommended in the 'CEOS Purge Alert Procedure'.

Description

“Data Purging” is the permanent and irrecoverable removal of EO data from an archive. The 'CEOS Purge Alert Procedure' helps minimize the loss of valuable EO data sets.

Guidelines

GUIDELINE 8.1 – Data purge alert procedure - (Level A)

Follow the “CEOS Purge Alert Procedure” when it has been determined that a certain data set cannot be preserved by the data holder any longer.

✓ CEOS EO Data Purge Alert Procedure [AD-4]

ANNEX A – GUIDELINES PRIORITY AND ASSOCIATION WITH RELATED DATA MANAGEMENT STANDARDS AND PRINCIPLES

The guidelines in this document have been categorized into levels A, B, and C. Adherence to Level A guidelines provides a basic level of security, integrity, and accessibility of the archived data. Level A guidelines, therefore, should be prioritized and implemented first. A more detailed description of the guideline levels is provided in Section 1.6 of this document and a summary is provided in Table 5 below.

An analysis of the level of compliance of the Preservation Guidelines to ISO 16363 - Audit and certification of trustworthy digital repositories and a comparison to the Data Management Principles of the Group on Earth Observations (GEO) - have been performed. For details see Section 1.7 of this document. The results of these comparisons are also provided in Table 5. For information, the GEO Data Management Principles (DMP), are listed in Table 4 below.

Discoverability:	
DMP-1	Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.
Accessibility:	
DMP-2	Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.
Usability:	
DMP-3	Data will be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.
DMP-4	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 extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.
DMP-5	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.
DMP-6	Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.
Preservation:	
DMP-7	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.

DMP-8	Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.
Curation:	
DMP-9	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.
DMP-10	Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

Table 4: GEO Data Management Principles

The following figure (Figure 1) shows a high level mapping between the Data Preservation Guidelines themes and the topics addressed in the GEO Data Management Principles.

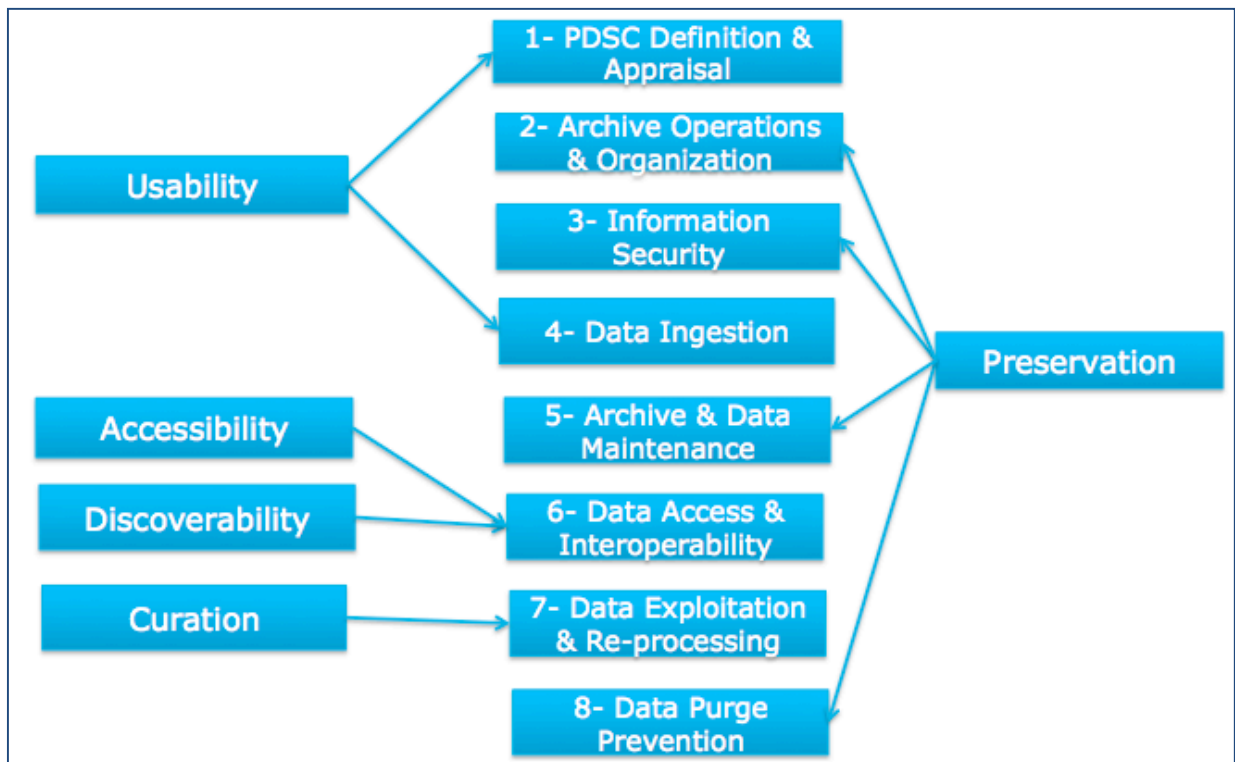


Figure 1: Mapping of the Preservation Guidelines themes (center) vs. the GEO Data Management Principles topics

Theme	Guideline	Level	ISO 16363		
			ISO Principles	Compliance (full, partial)	DMP Mapping
1. Preserved Data Set Content Definition & Appraisal	1.1	A	4.1.1	Full	DMP-4
	1.2	A	4.1.2	Full	DMP-5
	1.3	B	5.3 & 5.4	Full	DMP-4
	1.4	B	3.3.1	Partial	DMP-3
	1.5	B	4.2.1	Full	DMP-3
	1.6	B	4.2. (2-3)	Full	DMP-3
	1.7	B	3.3.1	Partial	DMP-3
	1.8	A	3.1.3	Partial	DMP-3
2. Archive Operations and organization	2.1	A			DMP-7
	2.2	A	3.3.2 & 3.3.3	3.3.2 (Full) & 3.3.3 (Partial but full in conjunction with other guidelines)	DMP-7
	2.3	A	3.3.3	Partial but full in conjunction with other guidelines	DMP-7
	2.4	A	Not covered		DMP-7
	2.5	A	3.2.1	Full	DMP-7
	2.6	B	3.4.2 & 3.5.2	Full	DMP-7
	2.7	A	4.16	Full	DMP-7
	2.8	A	4.2.4	Partial but full in conjunction with other guidelines	DMP-7
	2.9	C	4.2.4	Partial but full in conjunction with other guidelines	DMP-7
	2.10	A	4.3.1	Full	DMP-7
	2.11	C	4.3.3	Full	DMP-7
	2.12	A	4.3.2	Full	DMP-7
	2.13	C	Not covered		DMP-7
3. Information Security	3.1	B	5.2.2, 5.2.4 & 3.3.6	(5.2.2 & 5.2.4) Full & 3.3.6 (Partial)	DMP-7
	3.2	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7
	3.3	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7
	3.4	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7

	3.5	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7
4. Data ingestion	4.1	A	4.1.3,4.1.4,4.1.5, 4.1.6 & 4.1.7	Full	DMP-4
	4.2	A	4.2.6	Full	DMP-4; DMP-5
	4.3	A	Not covered		DMP-6
	4.4	A	3.3.5	Partial	DMP-6
	4.5	B	4.1.8	Full	DMP-6
5. Archive and Data maintenance	5.1	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-8
	5.2	B	4.2.5 & 4.3.1	Full	DMP-8
	5.3	C	Not covered		DMP-8
	5.4a	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.4b	B	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.4c	B	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.4d	C	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.5	A	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.6	B	5.2.1	Partial but full in conjunction with other guidelines	DMP-7; DMP-8
	5.7	B	4.2.9	Full	DMP-8
	5.8	A	Not covered		DMP-8
	5.9	A	Not covered		DMP-7
6. Data access and interoperability	6.1	A	Not covered		DMP-1; DMP-2
	6.2	A	Not covered		DMP-1; DMP-2
	6.3	B	Not covered		DMP-1; DMP-2
	6.4	A	4.6.2	Partial	DMP-1;
	6.5	A	4.6.1	Full	DMP-1; DMP-2
	6.6	B	Not covered		DMP-1; DMP-2

	6.7	B	Not covered		DMP-1; DMP-2
	6.8	C	Not covered		DMP-1; DMP-2
	6.9	C	Not covered		DMP-1
	6.10	B	4.6.2	Full	DMP-1
7. Data exploitation and reprocessing	7.1	A	4.2.5	Partial	DMP-9
	7.2	B	4.2.5	Partial	DMP-9
	7.3	B	4.2.4	Partial	DMP-9
	7.4	C	Not covered		DMP-9
	7.5	B	4.2.5	Partial	DMP-9
	7.6	A	4.2.5	Partial	DMP-9
	7.7	B	4.2.5	Partial	DMP-9
	7.8	C	4.2.5	Partial	DMP-9
	7.9	A	3.3.5	Full	DMP-9
	7.10	C	Not covered		DMP-9
	7.11	C	Not covered		DMP-9
	7.12	B	Not covered		DMP-10
8. Data Purge Prevention	8.1	A	3.1.2	Partial	DMP-7

Table 5: Guidelines priority levels and relationships