

Shared Collection Lifecycle Management Principles for Earth Observation Data

Version 0.2

September 2024

*Table of Contents*

Contents

[Introduction 3](#_Toc179835307)

[Purpose and scope of the document 3](#_Toc179835308)

[Project Members 4](#_Toc179835309)

[Applicable and Reference Documents 5](#_Toc179835310)

[Collection Lifecycle Management Principles 6](#_Toc179835311)

[Initialization Stage 6](#_Toc179835312)

[PRINCIPLE 1: Products should be consistent within Collections 6](#_Toc179835313)

[PRINCIPLE 2: Collection Upgrades should be deliberate, meaningful and well considered 7](#_Toc179835314)

[PRINCIPLE 3: At conception of a Collection Upgrade, cost storage, computation requirements, open-sourcing of production code and end-of-life preservation costs should be considered 7](#_Toc179835315)

[PRINCIPLE 4: Consider opportunities for interoperability and alignment with products in other organization’s collections. 7](#_Toc179835316)

[Detailed Definition Stage 7](#_Toc179835317)

[PRINCIPLE 5: Provisional products are part of the development process and their management and control are expected to differ from those of standard products within a collection 8](#_Toc179835318)

[Implementation Stage 8](#_Toc179835319)

[PRINCIPLE 6: Ensure Collection infrastructure providers maintain data integrity that enables disaster recovery 8](#_Toc179835320)

[PRINCIPLE 7: Confirm archive and disposal at end-of-life at publication stage 8](#_Toc179835321)

[PRINCIPLE 8: Do not delete or modify data without appropriate review, authorisation, and end user communication 8](#_Toc179835322)

[PRINCIPLE 9: Maintain accessible records of product files that have been removed or replaced within a collection 9](#_Toc179835323)

[PRINCIPLE 10: Consider approaches that authenticate replicas of collections 9](#_Toc179835324)

[Operations & Maintenance Stage 9](#_Toc179835325)

[PRINCIPLE 11: Maintain accessible metadata and documentation at the Dataset level for decommissioned Collections. 9](#_Toc179835326)

[PRINCIPLE 12: Communicate early and as required with stakeholders on pending decommission to the Collection 10](#_Toc179835327)

[PRINCIPLE 13: Source, ancillary, custom auxiliary data. algorithms and production software should be retained indefinitely 10](#_Toc179835328)

[PRINCIPLE 14: Obsolete collections should be disposed of in line with relevant entity and government regulations, polices or procedures 10](#_Toc179835329)

[PRINCIPLE 15: Only one version of a baseline data product should be discoverable within a collection 11](#_Toc179835330)

[PRINCIPLE 16: Ensure the survivability of the collection 11](#_Toc179835331)

[Appendix B: Terms of Reference 12](#_Toc179835332)

# Introduction

## Purpose and scope of the document

In an era when large volumes and varieties of Earth Observation (EO) data are proliferating across many different providers, it is important to document where fundamental approaches to the lifecycle management of EO collections broadly converge..

The Data Collection management highlights challenges including data integrity, data authenticity, data replica management (in the cloud and/or archived), data reproducibility and citation. A controlled Collections management enables collections (including metadata) to be properly handled and documented, and to become a resource for the user community.

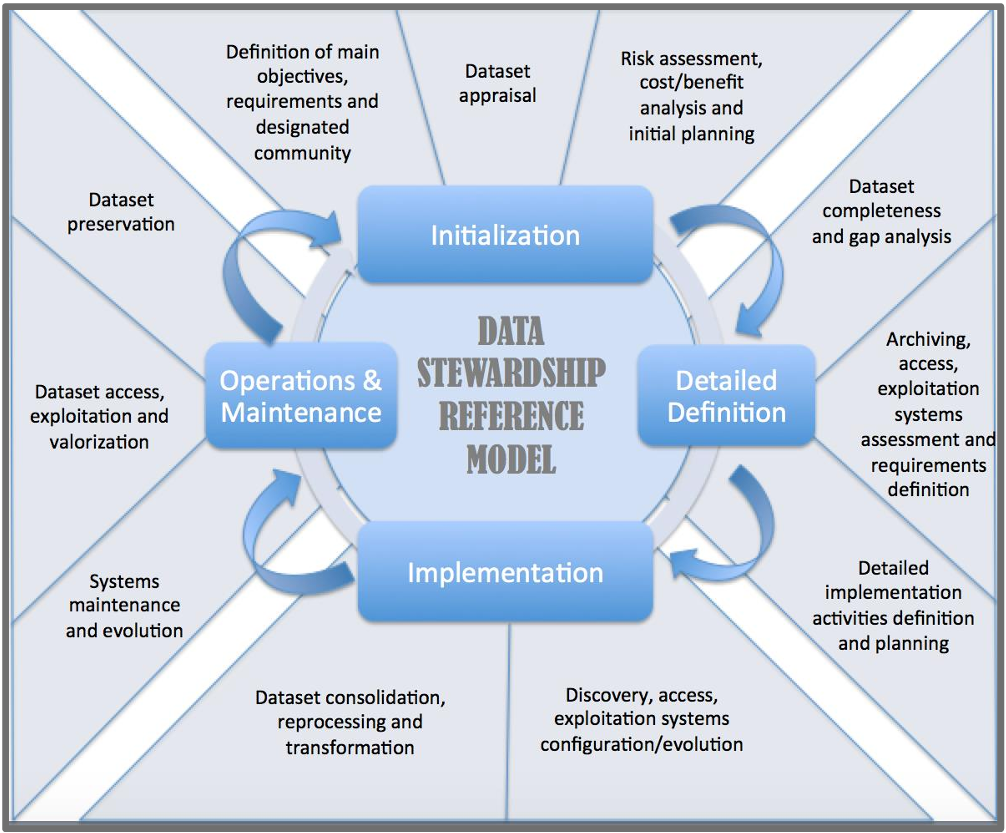
The benefits of Data collections include:

* + Consistent data accessibility, usability, and long-term preservation.
  + FAIR (Findable, Accessible, Interoperable, and Reproducible) scientific practices.
  + Facilitate the reanalysis of satellite measurements.

Collections are often generated through the reprocessing of the archive and therefore also referred to as “reprocessed datasets”.

The purpose of this document is to describe a set of core Earth Observation (EO) Collection lifecycle management principles shared by CEOS organisations

The principles were collaboratively derived in consideration of current agency practices and framed by a generalised collection lifecycle as represented by the Data collection Stewardship Model [AD 7] (figure 1). In the figure below, the major related activities are reported for each ECSS mission phase. The Data collection Stewardship process starts during the initialization phase and continues until the Operations and Maintenance phase. For each new data reprocessing the process needs to be restarted.



*Figure 1 Data collection Stewardship Model*

A significant body of work exists in the public domain relating to EO data management and preservation processes (2.3 Applicable and Reference Documents), but not always with collection lifecycle management as the animating concept. However, we sought to ensure general alignment of the shared collection lifecycle management principles with concepts contained in these references. Key terms of reference can and do vary amongst organizations and often pose impediments to shared understanding. Appendix B sources and clarifies the terms of reference used in this document.

## Project Members

This document has been developed with oversight from the CEOS Working Group on Information Systems and Services (WGISS), Data Stewardship Interest Group (DSIG).

The project team members are listed below.

|  |  |
| --- | --- |
| **Organisation** | **Representative** |
| USGS | Thomas Maiersperger |
| USGS | Steven Covington |
| Geoscience Australia | Simon Oliver |
| Starion for ESA | Iolanda Maggio |
| ESA | Mirko Albani |
| ESA | Damiano Guerrucci |
| DLR | Katrin Molch |
| NASA | Douglas Newman |

## Applicable and Reference Documents

|  |  |
| --- | --- |
| **ID** | **Resource** |
| [AD 1] | CEOS. 2015a. EO preserved dataset content. Retrieved from https://ceos.org/publications-key-documents/#best-practices-guidelines |
| [AD 2] | CEOS. 2015c. Long term preservation of earth observation space data: Preservation workflow. Retrieved from https://ceos.org/publications-key-documents/#best-practices-guidelines |
| [AD 3] | CEOS. 2016. EO data purge alert procedure. Retrieved from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/WGISS_DSIG_Data%20Purge%20Alert_WP.pdf> |
| [AD 4] | CEOS. 2017b. EO data glossary of acronyms and terms. Retrieved from <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| [AD 5] | CEOS. 2018. Measuring of earth observation data usage. Retrieved from <https://ceos.org/document_management/Working_Groups/WGISS/Documents/WGISS%20Best%20Practices/Measuring%20Earth%20Observation%20Data%20Usage%20-%20Best%20Practice.pdf> |
| [AD 6] | CEOS. 2019b. EO data preservation guidelines best practices. Retrieved from <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| [AD 7] | WGISS Data Stewardship Reference Model White Paper. Retrieved from https://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Data\_Stewardship/White\_Papers/WGISS\_Data%20Stewardship%20Reference%20Model%20White%20Paper.pdf |

Table 1. Applicable Documents

|  |  |
| --- | --- |
| **ID** | **Resource** |
| [RD 1] | Brobia, A., Voidrot, M.-F., Bye, B. L., and Maso, J., “GEO Data Management Principles: Updated Implementation Guidelines to increase the value of Earth Observation data”, vol. 2022, 2022. |
| [RD 2] | Mirko Albani & Iolanda Maggio (2020) Long time data series and data stewardship reference model, Big Earth Data, 4:4, 353-366, DOI: [10.1080/20964471.2020.1800893](https://doi.org/10.1080/20964471.2020.1800893) |

Table 2. Reference Documents

# Collection Lifecycle Management Principles

The principles and supporting rationale outlined in this section provide guidance to entities on recommended Collection Lifecycle Management practice from the perspective of CEOS data stewards. The principles are intended to be generally applicable, though the authors understand that some of the principles may not be able to be adhered to due to budgetary or political constraints

## Initialization Stage

### PRINCIPLE 1: Products should be consistent within Collections

Since a Collection represents a set of products that are processed using common specifications, products within a Collection are expected to be consistent. Minor revisions or updates to relatively small numbers of products within a Collection are expected in order to fix non-routine issues or errors that bring the products into a state of consistent quality. If updates or corrections are made to products, users should be able to reference the nature and magnitude of those changes and identify updated products.

### PRINCIPLE 2: Collection Upgrades should be deliberate, meaningful and well considered

A Collection Upgrade is characterized by a full reprocessing campaign to introduce major improvements or enhancements to all products. Changes are expected to break the consistency and continuity of product specifications between Collections. Therefore, major recalibration efforts, major algorithmic advances, and other changes to the structural specification of the Collection are held until upgrade time.

Decisions about starting a new Collection are of strategic nature and are made only infrequently at the higher level of the producing organisation. Decisions about the specification and timing of a Collection upgrade will depend on factors such as budget, producer priorities, user needs, advancements of scientific algorithms, and other technical considerations.

### PRINCIPLE 3: At conception of a Collection Upgrade, cost storage, computation requirements, open-sourcing of production code and end-of-life preservation costs should be considered

Initiation of a new collection should consider:

* estimates of storage requirements and growth rate over time (including periods of overlapping collection data persistence);
* computational resource required for back processing and ongoing production;
* end-of-lifecycle preservation options (archiving or disposal).

This information is needed to adequately estimate the cost of activities associated with the development, production, and dissemination of the proposed data product and to ensure that sufficient resources are provided for the project.

### PRINCIPLE 4: Consider opportunities for interoperability and alignment with products in other organization’s collections.

One of CEOS’s primary objectives is to optimise the benefits of space-based Earth observation through cooperation of CEOS Agencies in the development of compatible data products. Additionally, CEOS has the goal of encouraging complementarity and compatibility among space-based Earth observation system and the data received from them. The upgrade of a collection represents an infrequent opportunity to align product specifications to fulfil these objectives.

## Detailed Definition Stage

### PRINCIPLE 5: Provisional products are part of the development process and their management and control are expected to differ from those of standard products within a collection

Products with provisional status may be available in the public domain for research and development purposes (e.g., community assessment) and may not always lead to a final product. The nature of such products means they may be made available on-demand or in temporary data stores which are deleted at some point and are not archived. Provisional products may in fact change during their availability due to algorithm advancements or other factors during development. The provisional nature of these products must be communicated and caveated to users. Provisional products are normally superseded with the final version of products published into a given collection. In addition to publicly available provisional products, there may also be internal development product versions at various stages of completion but these would not be considered part of the Collection..

## Implementation Stage

### PRINCIPLE 6: Ensure Collection infrastructure providers maintain data integrity that enables disaster recovery

Consider the ability of supporting infrastructure to sustain the operational requirements of the Collection including: offsite copies, redundant storage / disk technology, business continuity, cyber security, disaster recovery and communication.

### PRINCIPLE 7: Confirm archive and disposal at end-of-life at publication stage

The decisions about archive or disposal at the end of the collection lifecycle should be reconfirmed and well communicated with sufficient lead time to enable a tactical response from users.

If disposal is the selected option, then it is expected that all relevant documentation and metadata required by organizational policy are preserved and made available to end users.

### PRINCIPLE 8: Do not delete or modify data without appropriate review, authorisation, and end user communication

Once formally published, final data variants cannot be deleted or modified without proper authorisation (with exception of minor bug fixes with no potential to impact users) and end user communication.

Bugs in final products: Sometimes a subset of the final product may be affected by subtle changes in input or auxiliary data that were not announced in advance, or due to interruptions in production . This may lead to the generation of faulty products which, when discovered, are deleted from the current collection and are replaced with the corrected products when available. Minor updates affecting relatively few individual data products within a collection are expected and corrections are expected. These updates do not "break the Collection" since they are not of the scope requiring a full reprocessing, or major version upgrade of a Collection.

New datasets may be added to a collection without the need for reprocessing a Collection. For example, as the result of addition of data from a new mission, or through publication of a new type of dataset.

### PRINCIPLE 9: Maintain accessible records of product files that have been removed or replaced within a collection

Records should be kept which allow users to identify affected products/ files and indicates whether they have been modified, removed and / or replaced.

### PRINCIPLE 10: Consider approaches that authenticate replicas of collections

Collection data may be replicated across multiple independent repositories (e.g. Physical Storage, Cloud Storage systems, etc.). Understanding how well a replica represents the original source collection is important in building user confidence. Ensuring users are able to confidently access authenticated copies of a collection is essential in providing user trust in a particular source of data. Different approaches could grant the data integrity and authenticity of the collections like the insertion of the PID in the product metadata, HASH code embedded in any replicated and disseminated product, digital signature (e.g. watermark).

## Operations & Maintenance Stage

### PRINCIPLE 11: Maintain accessible metadata and documentation at the Dataset level for decommissioned Collections.

There are two types of metadata records generated for products:

1. those related to collections, baseline and derived products or web services, and;
2. those related to a set of files comprising the individual product.

The higher-level metadata information (i.e., the first type) is generated and never deleted, even if the products they refer to are no longer available. Instead, they are marked as ‘deprecated’ or their status similarly noted. This way, old, decommissioned products can still be referenced meaningfully in historical scientific publications and online resources despite the data no longer being accessible.

The lower-level metadata records (very detailed) generated for each product are deleted at the time of product decommission (no archiving).

At the collection/product decommission stage, public access to the data is disabled. Although ongoing production may be discontinued, the data is preserved for a period of time in its original location.

Removal of public access means this data will no longer be referenced in catalogue applications and therefore, relevant indexes supporting those applications should be preserved (as snapshots in time). See also Principle 13 regarding obsolete collections.

### PRINCIPLE 12: Communicate early and as required with stakeholders on pending decommission to the Collection

Stakeholders should be consulted about the intention to decommission a collection and provided with an opportunity to feedback on potential impacts and preparedness in case changes in decommission timelines may be accommodated. Communication should include reference to the decision and whether the data will be archived and any period of availability during transition to the new collection.

### PRINCIPLE 13: Source, ancillary, custom auxiliary data. algorithms and production software should be retained indefinitely

Both source and auxiliary data are copies of data, often generated, and distributed by third parties. This data is typically of long-term strategic importance and hence, should retained and preserved indefinitely except when another agency has preservation responsibilities.

When no longer used in production, source and auxiliary data should be archived for long-term preservation (often on magnetic tape for backup).

Retention of these data enables reproduction of superseded versions of products (e.g. for legal purposes). It is worth noting that reproduction of old product versions may not always be possible. e.g. if the licences for the proprietary software used in the production expire or if the software / code used is no longer supported.

Auxiliary data is generally reusable across various collections and various versions of these collections. When newer substitutes are available, the old version is preserved and current version is maintained for production purposes (e.g., MODIS 6.0 was recently replaced with version 6.1).

### PRINCIPLE 14: Obsolete collections should be disposed of in line with relevant entity and government regulations, polices or procedures

Custodians of geoscientific data acquired on behalf of the government are responsible for managing the data through its lifecycle, including archiving.

Preserving obsolete collections in full may be cost-prohibitive due to the volume of data involved, hence old collection data and products are deleted when the upgraded collection data and products are published.

A representative example of a deleted collection can be retained for preservation purposes. See also Principle 10.

In deletion of a product version or entire collection, consideration should be given as to whether updated products will be a sufficient and acceptable substitute for the older version and/or whether the older version can be recreated from input and ancillary data and preserved algorithms (code, or the documented methods to produce the product).

Although the products can be legally disposed of, special consideration needs to be given as to whether this act will put an entity into disrepute. For example, products are used in “life and death” decisions (e.g., in support of emergency response efforts of other federal and state agencies) then the information pertaining to such events may need to be made available to post-event commissions of enquiry or similar legal requests if there was a loss of human life.

If product use may give rise to legal proceedings in the future then at least relevant information bits should be retained for a time. Also, if data was specifically provided for use during natural disasters, a copy of that specific data should be preserved by the supplying section (the group doing the specific work/application/study).

At the final stage collections/ products marked for disposal cannot be deleted before a formal authorisation to do so is obtained from the relevant authority. Only upon authorisation, should the orderly deletion of the data from all relevant locations take place.

### PRINCIPLE 15: Only one version of a baseline data product should be discoverable within a collection

Multiple versions of a product may exist at any point in time due to active maintenance processes. Therefore, at any time, there ideally should be only one version of baseline data available to users, per collection. However, there may be more than one version of derived product generated over time from a given baseline data.

The issue of archiving of baseline / unenhanced mission data is dealt with only at the time of collection upgrade. For individual products, the decision to decommission a version of a product or to distribute concurrent versions of the product is vested with the respective product owner.

Archiving of obsolete versions of products may be undertaken on a temporary basis but generally, it is avoided since newer versions are of better quality and fully replace their predecessors. See Principle 8 for additional guidance.

### PRINCIPLE 16: Ensure the survivability of the collection

Programs may rely on storage service providers (e.g., AWS, NARA) for backing up current collections. This means that, although there are system integrity preservation measures implemented by the vendors ensuring some data recovery capabilities, neither the responsible program nor entity may have an active backup procedure in place.

For example, at Australia’s National Computational Infrastructure (NCI) both the primary super-computer and scratch are lustre file systems where file metadata and file block data are stored on entirely separate storage hardware, with different redundancy levels. File block data is at least RAID6. The metadata is more highly redundant, typically some form of double mirrors. It means that the risk of data loss due to the hardware failure is limited as data can be restored. However, the risk of data loss due to accidental data deletion and/or corruption is not comprehensively addressed.

Tape storage used for archiving data, should have full backup (“two copies in two locations” principle applies).

# Appendix B: Terms of Reference

Note: CEOS source documents referenced in the table below are currently being reviewed

|  |  |  |
| --- | --- | --- |
| **Term** | **Definition** | **Source** |
| ancillary (data) | Data which are not obtained from the sensor itself (usually provided in the science telemetry) and have the primary purpose to serve the processing of instrument data. This can be divided into data referred to as spacecraft ‘engineering’, ‘core housekeeping’ or ‘subsystem’ data obtained from other parts of the platform and includes parameters such as orbit position and velocity, attitude and its range of change, time, temperatures, pressures, jet firings, water dumps, internally produced magnet fields, and other environmental measurements. Ancillary refers to data that exist purely to serve the data processing; auxiliary data, while helping the process, are also data sets in their own right. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| archive | The archive stores data products, guaranteeing their preservation for future use. This function includes all operations to identify, store and retrieve the data and ensure their integrity. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| auxiliary (data) | Data which enhance processing and utilization of remote sensing instrument data. The auxiliary data are not captured by the same data collection process as the instrument data. Auxiliary data include data collected by any other platform or process, preferably in georeferenced digital format. Examples are meteorological data received from ECWMF or NCEP. Auxiliary data help in data processing but are also data sets in their own right; ancillary refers to data that exist purely to serve the data processing (e.g., orbit position and velocity).  Auxiliary refers to any additional information which is required in production of an EO Collection . Examples include TLE (two-line element set) files to compute precise orbits of satellites, or water vapour estimates or BRDF MODIS (Bidirectional Reflectance Distribution Function) data for atmospheric corrections, also DEM (Digital Elevation Model) data for terrain corrections. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf>  Second paragraph from this document. |
| backup | Backup refers to a general practice of making copies of current baseline and derived products of current EO Collections, in specified time intervals, and storing those copies in a manner that will not affect the data if the primary distribution environment is compromised. The emphasis is on the ability to retrieve a copy of the entire EO Collection on demand if the original data is lost or compromised. | This document |
| baseline (data) | Baseline data are source data that has been processed to a common set of requirements and organised into a form that allows immediate analysis and interoperability through time and with other datasets. | This document |
| business continuity | Business continuity is an organization's readiness to continue functioning during times of disruption. | This document |
| catalogue *(noun)* | The Catalogue provides the discovery of information to the user on which EO data products can be obtained, i.e. a “Product Catalogue”. Products can be organized in collections with restricted access depending on product and user type. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| collection (general) | The ensemble of some products or auxiliary data having a common focus or theme or purpose (e.g., collection of land photos). | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| collection (specific) | A group of datasets that have been processed using a consistent radiometric calibration, geometric registration base, data / metadata format, and version of processing algorithms | This document |
| custodians/stewards | Data custodians are responsible for the safe custody, transport, storage of the data and implementation of business rules. | This document |
| data | Scientific or technical measurements, values calculated therefrom, observations, or facts that can be represented by numbers, tables, graphs, models, text, or symbols which are used as a basis for reasoning and further calculation. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| dataset | A logically meaningful grouping or collection of similar or related data. Data having all of the same characteristics (source or class of source, processing level, resolution, etc.) but different independent variable ranges and/or responding to a specific need are normally considered part of a single dataset. A data set is typically composed of products from several missions, gathered together to respond to the overall coverage or revisit requirements from a specific group of users. In the context of EO data preservation a data set consists of the data records and their associated knowledge (information, tools). See *collection*. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| dispose | To dispose of data is to permanently and irrecoverably remove all copies of an EO data set held by an organization. See *purge*. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| decommissioned | Products, datasets, and collections that have been superseded by new versions and are no longer available. Decommissioned products are no longer supported i.e corrections will not be made, new data will not be added, support is not available. | This document |
| deprecated | Products, datasets, and collections that have been superseded by a version and are still available, but whose use is discouraged since they will be decommissioned in the future. | This document |
| derivative (products) | Derivative products are additionally processed from baseline or unenhanced mission data and generally are either a compilation of time-series data (e.g., all time, annual, monthly or seasonal summary data, > Level 3) or thematic versions of each individual data tile of baseline data. Derived products are citable via a Digital Object Identifier (DOI). | This document |
| entity | A government or business organization that is formed to conduct business or represent the government of the day. | This document |
| final (version of product) | A product version that is not expected to be updated within a EO Collection. In its production, all of the prerequisite auxiliary data were input. | This document |
| granule | The smallest aggregation of data which is independently managed (i. e. described, inventoried, retrievable). Granules may be managed as logical granules and/or physical granules. | <https://www.earthdata.nasa.gov/learn/glossary> |
| index | Database entry linking to location of product. Indexed products are discoverable and visible in a catalogue. | This document |
| instrument data | 1) Data specifically associated with the instrument, either because it was generated by the instrument or included in data packets identified with that instrument. These data consist of instrument science and engineering data, and possibly ancillary data. Instrument engineering data is produced by engineering sensor(s) of an instrument, used either for operating the instrument or for processing the science data generated by the instrument. Instrument science data is produced by the science sensor(s) of an instrument, usually constituting the basic reason for existence of an instrument. (2) Data created by an instrument including scientific measurements and any engineering or ancillary data which may be included in the data packets. (3) Data produced and transmitted by the science and engineering sensors of an instrument, and, in the spacecraft environment, any additional data packaged with the instrument’s sensor data by virtue of services provided | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| Interim (version of product) | Temporary product destined to be superseded by subsequent versions of a product. | This document |
| Long Term Preservation | The act of maintaining information in a correct and independently understandable form over the long term ( i.e., a period of time long enough to be concerned with the impact which changing technologies, including support for media and data formats, and changing user communities will have on the information being held in a repository.) See also preservation. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| Master Data Record | In the context of EO data stewardship – the consolidated and quality checked dataset, which represents the result of applying the consolidation procedure. It is this master data record which should be used for preservation, dissemination, and any further processing / reprocessing. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| Metadata | Data about data, which provides an understanding of the content and utility of the dataset. Metadata may be used to select on data for a particular scientific investigation. Metadata is intended as information describing significant aspects of a resource (Earth Observation space data in this context). They are created for the purposes of data search, discovery and access management and may exist at various levels, typically from data collection through to the individual variables of each data file in a collection. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| NRT Near-Real-Time (data) | NRT Data are those that are available for use with a specified (small and application dependent) latency, which is typically 3 hours for meteorological applications. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| obsolete (datasets or products) | Obsolete products and datasets have been superseded by later versions of a product or dataset. | This document |
| preservation | Preservation covers all processes and operations on individual or multi-mission data sets for ensuring the technical and intellectual survival of Space Dataset and their metadata through time. It grants dataset integrity, its discoverability and accessibility, and facilitates its (re)-use in the long term. Preservation is one of the tasks of data curation. Examples are data record improvement and consolidation | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| Collection lifecycle management principle | High level guidance used within idealised EO collection lifecycle management | This document |
| product | The term is used in various Earth observation contexts – and with different meanings. Electronic data package distributable to users; content is derived from instrument data via processing involving ancillary and auxiliary data. Products may comprise metadata and browse images. A product may be part of a collection – a distinction useful for archiving and cataloguing purposes. The term product may be used to denote a product type, such as e.g. ENVISAT\_ASAR\_L1B\_PRI data. End users may distinguish between (input, "raw") data and products, i.e., the derived geophysical parameters. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| provisional | Products or datasets in a research and development phase and have not yet been published into a standardized collection. | This document |
| scene | Subset of an EO instrument acquisition data segment, cut by time (i.e., across-swath.) | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| source (data) | Data required for production of a EO Collection (e.g., unenhanced satellite mission data) data as well as further processed Level-1 data (precision geometrically and radiometrically corrected radiance at the sensor), or any other pre-processed format that suppliers are providing access to that meets program requirements. | This document |
| EO Space Data (CEOS term) see Instrument data | Earth Observations Data generated by spaceborne missions or instruments owned by public or private organizations. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| update (Collection) | An EO collection update makes minor revisions to a relatively small number of products within a dataset or collection for the purpose of correcting errors and maintaining consistent quality. | This document |
| upgrade (Collection) | An EO collection upgrade is a new and improved major version of all products and datasets within a collection due to comprehensive reprocessing. | This document |
| user | External person, institution or system that consumes user services (Data Access or Science and Service Exploitation Platform) provided by the payload data ground segment. | <https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf> |
| version (control) | Data version control is a method of working with datasets. It is similar to the version control systems used in traditional software development but is optimized to allow better processing of data and collaboration in the context of data analytics, research, and any other form of data analysis. Data version control may also include specific features and configurations designed to facilitate work with large datasets and data lakes.[1] | [Data version control - Wikipedia](https://en.wikipedia.org/wiki/Data_version_control) |