

CEOS

Working Group on Information Systems and Services
Data Stewardship Interest Group

EO Collection Citation Guidelines

CEOS
Data Stewardship Interest Group

Doc. Ref.: CEOS.WGISS.DSIG.EOCC
Date: July 2026
Issue: Version 1.0

Contents

EXECUTIVE SUMMARY.....	3
INTRODUCTION AND SCOPE.....	3
WHY DATA COLLECTION CITATION MATTERS.....	4
PRINCIPLES OF EO COLLECTION CITATION.....	4
EO COLLECTION CITATION MODEL.....	5
REQUIRED COMPONENTS OF A DATA CITATION.....	5
RECOMMENDED COMPONENTS OF A DATA CITATION.....	6
OPTIONAL COMPONENTS OF A DATA CITATION.....	6
DOMAIN-SPECIFIC COMPONENTS OF A DATA CITATION.....	6
ANNEX 1 - Recommended Citation Format.....	6
ANNEX 2 – Space Agency Metadata Survey.....	8
REFERENCES.....	9

EXECUTIVE SUMMARY

Earth Observation (EO) data collections constitute the foundation of scientific research, climate monitoring, environmental services, disaster management and numerous operational applications. As Earth Observation missions increasingly generate petabyte-scale archives that are continuously reprocessed, curated and enriched, the need for consistent and persistent citation of EO collections has become essential.

Data citation is no longer limited to individual collections. Scientific reproducibility, FAIR principles, long-term stewardship and proper attribution require that curated EO collections themselves become first-class citable digital objects.

This white paper proposes guidelines for the citation of Earth Observation collections, defining and suggesting the required, recommended, optional, and a few domain-specific components.

INTRODUCTION AND SCOPE

Data citation is the practice of providing a formal reference to a collection or other digital resource used to support research, analyses, or publications. It enables collections to be referenced in the same manner as traditional scholarly outputs, ensuring that the data underlying scientific results can be accurately identified, accessed, and acknowledged.

Despite the increasing availability of research data, collections are not yet cited as consistently as journal articles or other scientific publications. Research organisations therefore play a key role in promoting good data citation practices by establishing persistent links between published collections and the scientific literature that uses them.

A scientific publication should include sufficient data citation information to enable readers to identify, locate, and access the collections on which its findings are based. Proper data citation also acknowledges the creators and custodians of the data, supports transparency and reproducibility, facilitates data reuse, and enables the assessment of the provenance, quality, and long-term value of scientific data.

Citing data properly is essential for ensuring transparency, reproducibility, and giving appropriate credit to data creators.

Earth Observation programmes produce an increasing diversity of digital assets, including:

- satellite imagery
- calibrated products
- higher-level geophysical products
- ancillary information
- calibration and validation collections

- processing software
- documentation
- quality reports
- provenance information

These assets are commonly organized into curated collections representing coherent scientific products maintained throughout their lifecycle. The EO data collections evolve through:

- reprocessing campaigns
- algorithm improvements
- quality corrections
- metadata enrichment
- addition of historical observations
- migration to new archive infrastructures

Consequently, collections become living scientific assets whose identity must remain stable while allowing new versions to be cited independently.

WHY DATA COLLECTION CITATION MATTERS

Proper citation provides benefits to multiple stakeholders.

Scientific community • reproducibility • transparency • proper attribution • impact assessment	Data providers • recognition • usage metrics • stewardship visibility
Archive operators • traceability • lifecycle management • preservation planning	Funding agencies • measurable impact • investment justification • long-term preservation assessment

PRINCIPLES OF EO COLLECTION CITATION

The following principles should govern EO collection citation.

Principle 1 – Persistence: Every curated collection shall receive a persistent identifier (PID).

Principle 2 – Uniqueness: Each citation shall identify one and only one collection version.

Principle 3 – Resolution: Every PID shall resolve to a landing page.

Principle 4 – Accessibility: Landing pages shall remain available even after collections become offline.

Principle 5 – Versioning: Major collection updates shall generate new citable versions.

Principle 6 – Provenance: Citation metadata shall document the origin and evolution of the collection.

Principle 7 – Interoperability: Citation metadata shall comply with internationally adopted metadata standards.

EO COLLECTION CITATION MODEL

The EO Collection Citation Model defines a standardised approach for citing and referencing Earth Observation (EO) data collections. It establishes the information required to uniquely identify an EO collection and provides a consistent citation format that enables discovery, access, attribution, reproducibility, and long-term stewardship.

The model applies to all levels of EO data collections, from mission-wide archives to thematic, processing-level, sensor-specific, and user-defined collections. It supports the citation of both static and evolving collections by incorporating persistent identifiers, versioning, temporal validity, and provenance information. In the following paragraphs, the suggested required, recommended, optional, and domain-specific components are reported.

REQUIRED COMPONENTS OF A DATA CITATION

- **Author:** Name of individual or organisational entity (Creator or agency) responsible for the creation of the data. This could also be the rights holder.
- **Title of the data:** Complete title of the data.
- **Year of publication:** Year the data was first published or disseminated for the version of the data used.
- **Version:** Version or edition number associated with the data.

- **Data publisher:** Organisational entity (e.g., data repository or archive) that makes the collection available by preserving and/or disseminating the data.
- **Persistent identifier:** Unique electronic identifier used to locate and access the data (usually a digital object identifier, DOI, ARK, etc.). If available, the persistent identifier should be included as a link (e.g., <https://doi.org/xxxxxx>) and used instead of the URL or other Internet link.

RECOMMENDED COMPONENTS OF A DATA CITATION

- **Contributor:** Additional contributors such as project members
- **Resource Type:** Type of collection such as imagery, model output, or product
- **Description / Abstract:** Summary of collection contents
- **Keywords / Subjects:** Thematic keywords for discovery
- **Temporal Coverage:** Time period represented in collection
- **Temporal Resolution:** Time interval between observations
- **Data Format:** File format such as NetCDF, HDF, GeoTIFF
- **Access URL / Landing Page:** Webpage where collection is accessed
- **Download Link:** Direct link for collection download
- **Access Date:** Date collection was accessed by user
- **License:** Usage license for the collection
- **Contact Information:** Contact person or helpdesk
- **Metadata Creation Date:** Date metadata record was created
- **Metadata Standard:** Standard used (ISO 19115, DataCite, etc.)

OPTIONAL COMPONENTS OF A DATA CITATION

- **Alternative Title:** Alternative collection name
- **Language:** Language used in collection documentation
- **Geographic Coordinates:** Latitude/longitude bounding box
- **File Size:** Size of collection files
- **Usage Rights:** Conditions for data reuse
- **Funding Information:** Funding organization or grant
- **Project Name:** Project under which collection was created
- **Related Publications:** Research papers related to the collection
- **Related collections:** Links to associated collections

DOMAIN-SPECIFIC COMPONENTS OF A DATA CITATION

- **Instrument:** Sensor or instrument used for data acquisition
- **Satellite / Mission:** Space mission responsible for the data
- **Processing Level:** Processing stage (e.g., Level-1, Level-2)
- **Spatial Coverage:** Geographic region covered by collection
- **Spatial Resolution:** Spatial detail or grid resolution

ANNEX 1 - Recommended Citation Format

Data citations drawn from commonly used style manuals:

- ✓ APA: Cool, H. E. M., & Bell, M. (2011). Excavations at St Peter's Church, Barton-upon-Humber [Data set]. doi:10.5284/1000389
- ✓ Chicago: (Footnote) H. E. M. Cool and Mark Bell, Excavations at St Peter's Church, Barton-upon-Humber (accessed May 1, 2011), doi:10.5284/1000389.
- ✓ MLA: Cool, H. E. M., and Mark Bell. "Excavations at St Peter's Church, Barton-upon-Humber." Archaeology Data Service, 2001. Web. 1 May 2011. <<http://dx.doi.org/10.5284/1000389>>. Oxford: Cool, H. E. M. and Bell, M. (2011), Excavations at St Peter's Church, Barton-upon-Humber [collection] (York: Archaeology Data Service), doi: 10.5284/1000389

Citation formats suggested by three data repositories:

- ✓ PANGAEA: Willmes, S et al. (2009): Onset dates of annual snowmelt on Antarctic sea ice in 2007/2008. doi:10.1594/PANGAEA.701380
- ✓ Dryad: Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) Data from: The strength of phenotypic selection in natural populations. Dryad Digital Repository. doi:10.5061/dryad.166
- ✓ Dataverse: Frederico Girosi; Gary King, 2006, 'Cause of Death Data', <http://hdl.handle.net/1902.1/UOVMCPSWOL>
UNF:3:9JU+SmVyHgwRhAKclQ85Cg== IQSS Dataverse Network [Distributor] V3 [Version].
- ✓ The Publishing Network for Geoscientific and Environmental Data (PANGAEA) - open access library and data publisher for the earth and environmental sciences.
- ✓ Giesecke, T et al. (2013): Validated sample ages from 823 EPD sites. doi:10.1594/PANGAEA.804597, **Supplement to:** Giesecke, Thomas; Davis, Basil AS; Brewer, Simon; Finsinger, Walter; Wolters, Steffen; Blaauw, Maarten; de Beaulieu, Jacques-Louis; Binney, Heather; Fyfe, Ralph M; Gaillard-Lehmdahl, Marie-José; Gil-Romera, Graciela; van der Knaap, Pim Willem O; Kunes, Petr; Kühl, Norbert; van Leeuwen, Jaqueline F N; Leydet, Michelle; Lotter, André F; Ortu, Elena; Semmler, Malte; Bradshaw, Richard H W(2013): Towards mapping

the late Quaternary vegetation change of Europe. *Vegetation History and Archaeobotany*, 12 pp, doi:10.1007/s00334-012-0390-y

- ✓ Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) - for biochemical dynamics is one part of NASA's Earth Observing Systems Data and Information System (EOSDIS) managed by the Earth Science Data and Information System (ESDIS) project.
- ✓ Gu J.J., E.A. Smith, and H.J. Cooper. 2006. **LBA-ECO CD-07 GOES-8 L4** Gridded Surface Radiation and Rain Rate for Amazonia: 1999. Data Set. Available online [<http://www.daac.ornl.gov>] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/831 **Data and Information System (ESDIS) project.**

ANNEX 2 – Space Agency Metadata Survey

Metadata Field	NASA (USA)	ESA (Europe)	JAXA (Japan)	ISRO (India)	CSA (Canada)	CNES (France)	DLR (Germany)	Common Standard (DataCite/ISO/CEOS)	Description
Instrument	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Domain-specific	Sensor or instrument used for data acquisition
Satellite / Mission	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Domain-specific	Space mission responsible for the data
Alternative Title	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Optional	Alternative dataset name
Language	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Optional	Language used in dataset documentation
Geographic Coordinates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Optional	Latitude/longitude bounding box
File Size	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Optional	Size of dataset files
Usage Rights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Optional	Conditions for data reuse
Funding Information	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Sometimes	Optional	Funding organization or grant
Project Name	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Optional	Project under which dataset was created
Related Publications	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Optional	Research papers related to the dataset
Related Datasets	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Optional	Links to associated datasets
Contributor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Additional contributors such as project members
Version	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Dataset version number
Resource Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Type of dataset such as imagery, model output, or product
Description / Abstract	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Summary of dataset contents
Keywords / Subjects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Thematic keywords for discovery
Spatial Coverage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Geographic region covered by dataset
Spatial Resolution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Spatial detail or grid resolution
Temporal Coverage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Time period represented in dataset
Temporal Resolution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Time interval between observations
Processing Level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Processing stage (e.g., Level 1, Level 2)
Data Format	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	File format such as NetCDF, HDF, GeoTIFF
Access URL / Landing Page	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Webpage where dataset is accessed
Download Link	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Direct link for dataset download
Access Date	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Date dataset was accessed by user
License	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Usage license for the dataset
Contact Information	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Contact person or helpdesk
Metadata Creation Date	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Date metadata record was created
Metadata Standard	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Recommended	Standard used (ISO 19115, DataCite, etc.)
Dataset Identifier (DOI)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Required	Persistent identifier for the dataset
Dataset Title	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Required	Name of the dataset
Creator / Author	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Required	Person or organization responsible for dataset creation
Publisher / Repository	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Required	Organization distributing the dataset
Publication / Release Date	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Required	Date dataset was made publicly available

REFERENCES

- FORCE11. *Joint Declaration of Data Citation Principles*.
- DataCite. *Metadata Schema Documentation*.
- CEOS WGISS. *Shared Collection Lifecycle Management Principles for Earth Observation Data*.
- CEOS WGISS. *Data Stewardship Reference Model*.
- W3C. *PROV Data Model (PROV-DM)*.
- Wilkinson, M. D., et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 3, 160018.
- Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS)* (Blue Book 650.0).