CEOS Interoperability Handbook

Enhancing Interoperability of data and services between different Earth Observation stakeholders.

Table of Contents

Introduction

During the last three decades, **Committee for Earth Observation Satellites (CEOS)** has significantly contributed to the advancement of space-based Earth observation (EO) community efforts. CEOS Agencies communicate, collaborate, and exchange information on EO activities, spurring useful partnerships. CEOS Agencies work together to launch multi-agency collaborative missions, and such cooperative efforts have highly benefited users all around the world. CEOS also provides an established means of communicating with external organizations, enabling CEOS to understand and act upon these organizations' EO needs and requirements.

Interoperability of data and services in the EO domain is essential due to the importance of EO data in a wide range of applications, including agriculture, climate monitoring, disaster management, and urban planning, as well as the increasing cooperation between public and commercial space actors worldwide. In order to streamline such collaborative efforts a need for guidelines on interoperability for data and services was strongly felt and the **CEOS Interoperability Handbook 1.1** was developed and published in 2008. With the increase in number of EO satellites and associated complex sensors providing ever growing volumes of data catering to complex user requirements, and over fifteen years since the first version, it is time review the interoperability of data services in the current context and development of the **CEOS Interoperability Handbook 2.0** was proposed.

The **CEOS Interoperability Framework** was developed to guide the evolution of this handbook for EO data services. Different entities within CEOS contributed towards development of this Handbook with CEOS **Working Group on Information Systems and Services (WGISS)** as a lead for overall coordination and development.

Purpose

This handbook will help data providers increase the interoperability of their data and services with those of other organizations, agencies, and countries by proposing standards and best practices in terminology, structure, formats, metadata, quality and policy. This handbook will help to avoid duplication of efforts, reduce costs, and improve the interoperability of EO data and services.

Interoperability also supports the integration of global EO data with other data sources, such as geo-statistics, model data, and forecasts. By combining these sources, users can gain a more comprehensive understanding of various phenomena, such as natural disasters, land use, and climate change at global level.

Furthermore, adaptation of interoperability recommendations as discussed in this handbook will facilitate data sharing among the scientific community, governments, international organizations and other stakeholders. It can assist in creating a common understanding among users with different backgrounds, interests, and needs. Finally, interoperability goes beyond data sharing, as it can foster collaboration and innovation by enabling the development of new applications, tools, and services that leverage EO data. Such advances are critical for addressing emerging challenges, such as the impacts of climate change, natural disasters, and food security.

Scope

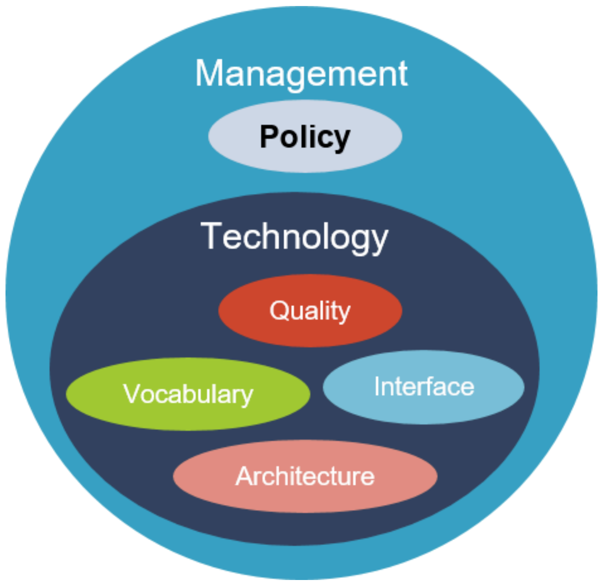
CEOS Interoperability Handbook Version 2.0 intends to provide guidance to the organizations for **development of Interoperable Data and Services** and help them in **measuring their maturity level**.

Audience

This Handbook is intended to be used as a guide by different **Space Agencies, New Space Startups** and **Commercial Data and Service Providers** to improve interoperability of data and services.

Interoperability Framework

In order to guide and structure the overall Interoperability related activities within CEOS, a framework was proposed and developed. Five areas have been identified as critical 'factors' needed to move towards greater interoperability of EO data and services.



Each factor has several components and are summarized as below.

Factor	Description
Vocabulary (Semantics)	The (narrow) semantic aspect refers to the naming and defining terms and expression. It includes developing, harmonizing, and maintaining vocabularies, concepts, and schemata supporting provision, exchange, and analysis of data, information, and knowledge regarding Earth observation. It ensures that words and language are understood in the same way by all communicating parties
Architecture	Architecture describes the organizational structure of concepts, processes, and assets, including data. It comprises of the structural aspects of models and standards that govern the collection, storage, archiving, documentation and publication of data
Interface (Accessibility)	Data exchange protocols and application interfaces, from a consumption or user perspective. These provide the means necessary

	to search for collections, find and access data and information contained in those collections
Quality	Indicators (parameters, metrics, etc.) for informing users of the trustworthiness (accuracy, uncertainty, consistency, etc.) of the data provided (measurands, measurements, observations, etc.)
Policy	Legal frameworks, policies, rules, and strategies regulating the relation between the different stakeholders.

The factors together intend to cover all aspects that play a role in enabling interoperability of data and services. Design criteria for the factors are cohesion, independence, and modularity. Cohesion (internally) is accomplished by grouping them according to specific knowledge and competences required to tackle the respective issues. This 'separation of concerns' known from system engineering should ensure that the individual aspects are addressed in the most adequate way by experts of the respective fields without having to deal with the complexity of the entire system all the time. Independence means that specific (fundamental) aspects (e.g., data formats) should be proprietary to one factor. In this way parallel or duplicate solution development can be avoided. Modularity must ensure that all factors together build a functional ensemble in which it is clear which covers what and how they all work together.

The interoperability factors collate certain, expected-to-be related, competencies that should be considered whenever interoperability is desired or required in a project or process. Most CEOS activities are targeted at enhancing interoperability and thus they all should scrutinize their work with respect to the framework factors.

Vocabulary (Semantics)

Interoperability relies on the ability of diverse entities to communicate and exchange information seamlessly. At its core, semantics and vocabularies play a fundamental role in ensuring that data, messages, and services are not only transmitted but also correctly understood across different stakeholders. Without a shared understanding of terms, concepts, and relationships, interoperability remains limited, hampered by inconsistencies, misunderstandings, and other integration challenges.

Semantics deal with general aspects of meaning and relationships between terms and concepts in a domain, ensuring that communication is structured and interpreted consistently. As an important part of such communication, vocabularies, including thesauri, glossaries, terminologies, ontologies, taxonomies, and controlled vocabularies, provide standardized definitions that facilitate common understanding. In the context of geospatial interoperability, standardized vocabularies enable diverse entities to describe, classify, and relate data and services in a way that is human and machine-readable and reusable across the whole domain.

This section highlights the essential role of semantics and vocabularies in the CEOS Interoperability Framework. It outlines key standards, best practices for implementing semantic interoperability, and methods for aligning domain-specific vocabularies. By establishing a shared semantic foundation, stakeholders can improve data exchange, integration, service compatibility, and automated reasoning, ensuring more effective collaboration in an increasingly interconnected digital geospatial ecosystem.

ID	Semantic Recommendations
SEM#1	Terms and definitions should be collected into an open Earth observation thesaurus, such as that provided by KCEO/CEOS through GitHub.
SEM#2	Capability should be provided to enable public comment and discussion on existing and new terms and definitions.
SEM#3	Enable version control and change management at the individual term level and link to historical and alternative definitions.
SEM#4	Use of project or document specific vocabularies should be discouraged e.g., in the form of ‘terms and definitions’ chapters. Source (via weblink), maintain, and develop all terms that serve or might serve in more than one context in the online, shared repository.

SEM#5	Community members should promote the common thesaurus, including through ISO/TC 211, OGC, WMO, GEO and other stakeholders in Earth System Sciences, to strive for domain wide adoption.
SEM#6	Common online repositories for abbreviations and acronyms should be used. Agreed metadata fields with unified and binding lists of options should be included. Keywords from controlled vocabularies that allow lookup of keyword information via Linked Data principles, e.g., HTTP URI dereferencing or SPARQL interfaces are preferred. The use of GCMD controlled keywords is encouraged.

Of central importance for increasing the interoperability within CEOS and across the entire geospatial domain will be a more harmonised and structured terminology. Providers and users of EO data and services will largely benefit if the definition and interpretation of terms is no longer renegotiated and amended each time projects are started, or new documents are drafted. A key finding of the terminology task force was that vocabularies should be developed as much as possible in an open and participatory way across the whole domain they are intended for. One of the main lessons learned by the CEOS terminology task force was that usability and acceptance of unified vocabularies will largely depend on consistent and comprehensive principles shared by all stakeholders and guiding their development.

ID	**Thesaurus Recommendations**
THES#1	The terms used in the thesaurus should be consistent and divided into classes such as Base, Core, Controversial and High Impact. The ‘Base Terms’ should have cross community agreement and should not have circular or ambiguous definitions. The ‘Core Term’ should be using the ‘Base Term’ consistently and can be allowed to have minor tweaks with approval from the identified committee. The ‘Controversial Term’ should have qualifiers attached to them with links to discussions, which led to the association of the qualifier. The ‘High Impact Term’ should be approved by a specialist committee and should be linked to a document providing details of the term.
THES#2	The definition of a term may not contain the term itself nor other circular definitions (e.g., where term A is defined using term B and term B is defined using term A). A clear set of base terms should be used.
THES#3	The terms used in the thesaurus should have clear and mappable relationships with other terms (parent, sibling, child). Overlaps between terms

	that are supposed to delineate more generic concepts (siblings) should be avoided or minimized.
THES#4	Definitions have to be kept unambiguous and short, and written in a form such that they can replace the term in a sentence.
THES#5	Explanations should be given in a separate 'Notes' sections, and Examples in a separate 'Examples' section. Both complement the definition, and should not be included as part of the main definition.
THES#6	Every definition should have an accompanying 'Sources' section, where all source documents are listed, wherever possible as weblinks.
THES#7	Thesaurus terms should be version controlled at the individual term level.
THES#8	Where a term is deemed 'controversial' then contradictory definitions can be provided, but only with clear links to alternative definitions and explanations as to what context a term is used in.

Architecture

Architecture plays a very important role in enabling interoperability. It describes the organizational structure of concepts, processes, and assets, including data and workflows. It comprises structural aspects of models and standards that govern the collection management, archiving, storage, documentation and publishing of data, and is the basis on which the interoperability of data and services is built on.

The Architecture factor has been divided into the following sections:

1) Preservation Architecture
2) Data and Metadata Architecture
3) Publishing Architecture

Preservation Architecture

The following list of recommendations describes the elements contributing to archive interoperability.

The primary purpose of data archiving is to preserve data over time. Preserving data over time consists in holding data in repositories in a way that enables data to be managed and accessed now and in the future. Data archiving is a complex, long-term process, with possibly many partners, including data providers supplying data to the archive, data users willing to use the archive, archive managers organizing the archive and other archives with which interoperability may be sought. Data management and archiving should consider not just the storage of data, but also the access and usage patterns of data.

ID	Recommendations
DPRES#1	Archival systems should comply with the Reference Model for an Open Archival Information System (OAIS) and with the forthcoming "OAIS-Interoperability Framework" to facilitate interoperability between archives.
DPRES#2	Data should be appraised and properly documented before ingestion in the archives following the CEOS Data Appraisal Procedure.
DPRES#3	Data and associated information should be ingested, archived and preserved following internationally recognised standards and best practices (e.g., those

	produced by WGISS and Producer-Archive Interface Methodology Abstract Standard) with any tailoring documented.
DPRES#4	Periodically perform archival system/media upgrade to the most adequate proven technology to ensure data and information long term preservation. Perform migration, with an integrity check, of archived data from old to new systems.
DPRES#5	Archive and preserve the information, code and software needed to handle the archived data, following the CEOS guidelines.
DPRES#6	When performing archived data and information repackaging and/or reformatting, for example to comply with new standard formats and/or exchange formats, properly document changes made to the archived data and ensure data integrity.
DPRES#7	Periodically verify the integrity of the archive collection/content through integrity check on a representative set of the archived data.
DPRES#8	Manage evolution of archived data collections according to the Shared Collection Lifecycle Management Principles for EO Data best practice.
DPRES#9	Keep archives equipment (hardware and software) up-to-date and in conformance with vendor recommendations to preserve data and associated information integrity and facilitate interoperability between archives.

Data and Metadata Architecture

This section covers the core recommendations for collection management functions including data production, management, packaging and documentation.

ID	**Recommendations**
DATA#1	CEOS-ARD Framework should be used as a starting point for development of Analysis Ready Data.
DATA#2	CEOS-ARD Product Family Specifications (PFS) should be used for development and assessment of ARD products, including both self-assessments and peer review.

DATA#4	The ISO 19115 series of standards (or similar) should be used to produce geospatial metadata.
DATA#5	A Collection of data should have all granules packaged consistently and produced with consistent quality.
DATA#6	Collection-specific metadata formats may be used, but packaging must also include STAC documents at the Collection and Granule/Item level. Refer to the CEOS EO collection and granule discovery best practices with STAC.
DATA#7	Checksums for all files in a packaged granule should be available, to ensure integrity.
DATA#8	Where pixel-level metadata is avaliable, such as scene quality masks, these should be clearly documented with a reference to lookup tables.
DATA#9	File names and folder or path structures should be consistent and include appropriate information to distinguish the specific granule. This could include the platform, time and date of aquisition, band(s), and product version.
DATA#10	Assign a Persistent Identifier to data archived and published to users and ensure the availability of all associated information in the relevant Landing Page following the CEOS Persistent Identifiers Best Practice.
DATA#11	The CEOS supported Open Data Cube family of software can be taken as a reference Datacube implementation.

Publishing Architecture

Publishing recommendations involve the final stage in making data accessible to external organizations or individuals. These recommendations are aimed at facilitating both access to data as well as maintaining a replica of part or all of a collection of data, including to be used when managing data for interoperability on the cloud.

ID	Recommendations
PUBLISH#1	Each data collection that is published as a publicly-accessible product should include a public granule-level notification including for when a granule is added, updated and deleted/archived. This supports management and maintenance of replicas collections.

PUBLISH#2	A collection should have a full listing of all available granules in a standard format, preferably cloud optimized. For example, STAC-geoparquet is used by some providers.
PUBLISH#3	Granule data stored in the cloud should be accessible in cloud-optimized formats, e.g., Zarr or Cloud-Optimized GeoTIFF (COG).
PUBLISH#4	Granules should not be zipped when stored in the cloud, so that cloud optimized data formats can be leveraged.
PUBLISH#5	Where possible, cloud providers' standard interfaces should be used in preference to self-developed solutions, enabling interoperability of tools that work with that cloud provider.

Interface (Accessibility)

Interfaces allow diversified resources within and across the organization to seamlessly communicate, discover and exchange data. Interfaces are realized in the form of services and follow standards. Interfaces enable data users to have easy and efficient ways of discovering and accessing data and associated services through the exploitation of standard protocols and the harmonizing of search and data retrieval processes

Data Discovery

ID	Recommendations
DISC#1	Collection and granule discovery interfaces should comply with the CEOS STAC Collection and Granule Discovery Best Practices (preferred) or CEOS OpenSearch Best Practices.
DISC#2	Service and tool discovery interfaces should comply with CEOS Service Discovery Best Practice.
DISC#3	Collection and granule metadata obtained via the discovery interfaces should advertise the existence of the corresponding file-level online data access and subfile or pixel-based access services and endpoints (e.g., OGC WCS, WMTS, WCPS, OGC API Maps, OGC API Tiles, etc.).
DISC#4	Granule metadata obtained via the discovery interfaces should include the online data access URL to the granule (in full resolution) and to a low resolution representation (i.e., quicklook or thumbnail). The low resolution representation should be provided in Web-friendly format, e.g. JPEG or PNG, and may be a static file or an OGC WMS/WMTS or API Maps/Tiles response.
DISC#5	Discovery interfaces should be accessible and return responses without requiring authentication.
DISC#6	Collection and granule metadata obtained via the discovery interfaces should advertise the existence of the corresponding authentication endpoint for human and machine access to the data (if required).

DISC#7	Resource metadata including keywords should link each keyword to its URI and to the appropriate thesaurus (i.e., controlled vocabularies).
DISC#8	Keywords from controlled vocabularies that allow lookup of keyword information via Linked Data principles, e.g., HTTP URI dereferencing or SPARQL interfaces are preferred. The use of GCMD controlled keywords is encouraged.
DISC#9	Resource metadata should contain the persistent identifier (e.g., DOI) of the corresponding resource.
DISC#10	Collection metadata should refer to the level of maturity with respect to the WGISS Data Management and Stewardship Maturity Matrix.
DISC#11	For facilitating discovery and access, data shall be organised in collections according to the principles outlined in the WGISS Data Collections Management Practices White Paper.

Data Access

ID	Recommendations
DACC#1	Granule data stored in the cloud should be accessible directly via a web-based protocol, for example the S3 (Simple Storage Service) and HTTP(S).
DACC#2	Data access should support file-level access and subfile or pixel-based access. Data download interfaces over HTTPS should support `Range Requests` to allow clients to request a portion of a file. Typical use case: access to a portion of a Cloud-Optimized GeoTIFF (COG) file.
DACC#3	In case a granule consists of many individual assets (files), it shall be possible to access each asset individually and it is recommended to provide access to all subcomponents of a granule with a single request.

Authentication and Authorization

ID	Recommendations
AUTH#1	Authorization should be available at a file level for both human and machine-to-machine access.
AUTH#2	Authentication interfaces should comply with open standards, such as the OpenID Connect protocol.
AUTH#3	HTTPS requests for data access that require authorisation will support well known methods for both human and machine-to-machine interface, such as those specified in the OpenAPI 3.0.

Quality

Quality informs users of the trustworthiness of Earth observation data and products. Multiple Calibration and Validation (Cal/Val) groups/venues exist as forums for the exchange of information about understanding, expressing, and improving data quality, along with influencing the interoperability between multiple datasets and products.

Calibration and Validation

Calibration is the process of quantitatively defining a system's response to known and controlled signal inputs. Validation, on the other hand, is the process of assessing, by independent means, the quality of the data products derived from those system outputs.

Cal/Val Recommendations

ID	Recommendations
CALVAL#1	Data providers should engage and participate in community calibration/validation groups, such as CEOS WGCV (and its subgroups), WMO GSICS, JACIE and VH-RODA.
CALVAL#2	The Measurand and Uncertainty of stated values should be included within all products, as they are key to communicating and understanding data quality.
CALVAL#3	All products should have associated quality indicators, traceable to reference standards to allows users to assess usability of the data for their applications.
CALVAL#4	Post-launch, Level-1 products should be calibrated using CEOS Fiducial Reference Measurements (CEOS-FRM).
CALVAL#5	Community endorsed Cal/Val sites and reference networks should be used for satellite cross-comparison, such as CEOS Cal/Val sites, RadCalNet and SARCalNet.

CALVAL#6	The Quality Assurance Framework for Earth Observation QA4EO developed by Group on Earth Observations (GEO) and endorsed by CEOS should be followed to enable interoperability and quality assessment of earth observation data.
CALVAL#7	The ESA/NASA/USGS Earth Data Assessment Project (EDAP) should be used for reporting metrics related to quality.
CALVAL#8	The Joint Agency Commercial Imagery Evaluation (JACIE) Best Practices document should be used as a guideline for performing standard calibration and validation activities.
CALVAL#9	CEOS CAL/VAL portal should be used as a reference site for accessing agreed good practices and CAL/VAL protocols for interoperability for Earth observation calibration and validation activities.

Policy

Policy is a statement of intent, which provides guidance for implementation of processes and procedures in an organization. Policies can be at organization/local level or at the government/center level and may have legal bindings. The government level policies take precedence over the organization policies. Institutional mechanism is required to ensure compliance of policies in an organization. The policies are the guiding document for decision making processes in an organization.

Policy is one of the most important factors and forms the basis of interoperability. Following are interoperability recommendations for the policy factor.

ID	Recommendations
POL#1	Data providers should participate and engage in relevant community groups/events, such as CEOS, GEO and CGMS, and their respective working groups. Interoperability requires collaboration and coordination between all actors within the sector.
POL#2	Identify policies in your organization/country related to data and services and conduct periodic check/audit for compliance to these policies. Identify policies which may be barriers to interoperability of data and services and flag them for resolution. Ensure the policies are clearly communicated to stakeholders.
POL#3	**EO Capabilities:** Publish and periodically update information about present and planned Earth observation Satellites in online databases, preferably the CEOS MIM Database. This will help in planning and overall coordination among different EO stakeholders.
POL#4	**Open Standards and Specifications:** Ensure your organizations implement open standards and specifications such as those published by the OGC for data and services. Drafting of new specifications for data formats, metadata formats and service APIs should be preferably done along with standards organisations, or developed by the open source community.
POL#5	**Open Data:** Organizations should ensure that Earth observation data is discoverable, accessible and proactively made freely available for use, reuse and redistribution to users in human and machine readable form.

POL#6	**Open Source Software:** Where possible, share software applications as open source software, enabling others to use the same tools as are used internally to process or transform data products or to demonstrate the use of standards to access your data and services. An example of a preferred license is Apache 2.0.
POL#7	**Open Science:** Promote the concept of open science for collaborative development. Open science ensures availability of the state-of-the-art algorithms and software providing consistent products from different data providers and supporting reproducibility.
POL#8	**Data Licensing:** Organizations sharing open and unrestricted data should license the data using an open source license, consistent with their organisation's policy. A Custom license can restrict access for users. The GEO data licensing Guidance can be referenced for examples, including Creative Commons Zero 1.0 Universal Public Domain Dedication (CCo), Open Data Commons Public Domain Dedication and License (PDDL) v1.0, or Creative Commons Attribution 4.0 International (CC BY 4.0). CC BY 4.0 is preferred.
POL#9	**Data Procurement from third party:** Organizations planning to procure/outsource Earth Observation data, to possible extent should ensure that the data complies with CEOS recommendations, including those outlined in this handbook.
POL#10	**Data preservation:** Organizations should ensure that Earth observation data is archived and preserved according to CEOS best practices.
POL#11	**Purge Alert:** Organisations should use the purge alert service provided by CEOS WGISS before data and information removal from archives.
POL#12	Apply the FAIR principles to pursue data and metadata interoperability