CEOS LSI-VC-20

CEOS-ARD Unified Modeling Language (UML) diagram

Sumit Sen
Siri Jodha Khalsa

Agenda Item 3.3
LSI-VC-20
Reston, VA, USA
14-16 September, 2026

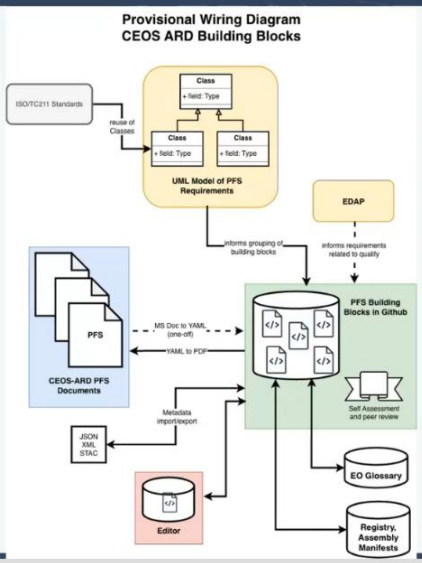
Committee on Earth Observation Satellites

CONTEXT

From Conceptual Blocks to Formalized PFS: The CEOS-ARD Metadata Story

Aligning ISO 19115-1/19115-2 standards with CEOS ARD Building Blocks

UML model of PFS Requirements	Based on ISO 19115 and extensions
Reusable building blocks	PFS building blocks in Github
Compliance and crosswalks	Documents to Github, Metadata imports/exports
Editing / Compliance	Metadata in registries and manifests



Why Formalization Matters

The core challenge isn't a lack of models, but the semantic distinction between terms.

Term	Ambiguity (Severity)	ISO Support (Sufficiency)	Clarity Priority
Band	Medium	High	Low
Boundary	Low	High	Very Low
Resolution	Low	Medium	Low
Pixel	High	Medium	High
Geolocation	Medium	Medium	Medium
Georeferencing	Very High	High	Very High
Georectification	Very High	High	Very High

The Complete Information Flow

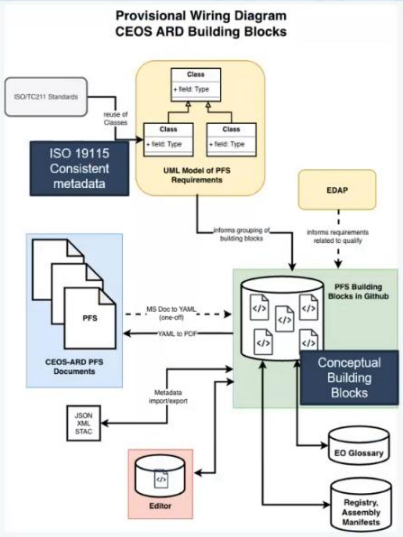
Solve ambiguity by linking conceptual building blocks to formal ISO metadata via a rigorous information flow.

ISO/TC211 Standards → Informs the UML Model of PFS Requirements

UML Model → Informs the grouping of PFS Building Blocks in Github

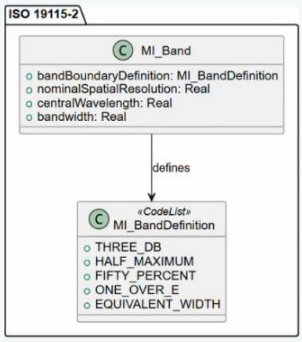
PFS Documents ↔ PFS Building Blocks (MS Doc to YAML → YAML to PDF)

All → Editing/Compliance Checking → Metadata Export (JSON/XML/STAC)



Defining the Core Terms (The Glossary Approach)

A standards reference must describe the nature of the relationship without implying formal conformity or endorsement.

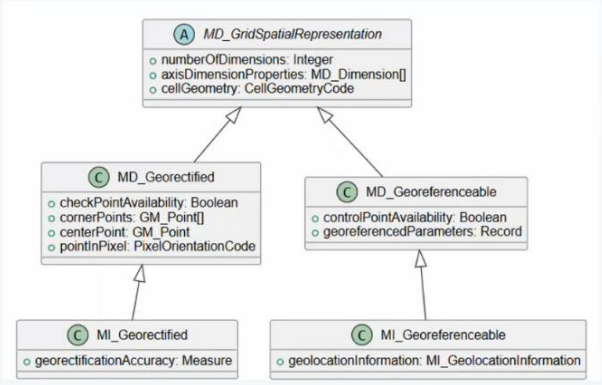


Band: In Earth Observation, a Band refers to a specific range of the electromagnetic spectrum recorded by a sensor. The semantics go beyond a simple number. It encompasses the spectral properties used to define it, which are crucial for radiometric interoperability. [Libby: Part of the electromagnetic spectrum of specific wavelengths. In Remote Sensing, usually described by central wavelength and bandwidth]

ISO 19115-2 provides the *MI_Band* Definition class, which is a CodeList that formally defines the criterion for specifying a band's spectral limits.

The key is not just that a band exists, but how its boundaries are defined. This is critical for comparing data from different sensors. For example, FWHM is a common definition, but CEOS ARD PFS often require the full Spectral Response Function (SRF) for better accuracy.

The MI_BandDefinition code list provides a standardized way to specify the criterion (e.g., THREE_DB for FWHM, EQUIVALENT_WIDTH) for defining maximum and minimum wavelengths. MI_Band class then uses this definition to record the band's properties, including its central wavelength and resolution, which is essential for meeting PFS requirements



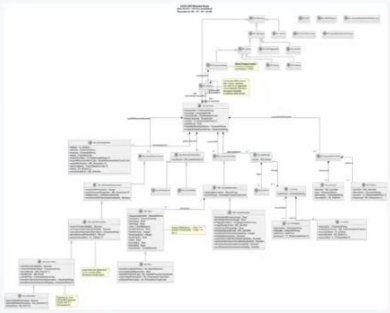
Boundary: This is the geographic extent of the dataset (e.g., bounding box). It is a high-level concept defined in ISO 19115 by classes like EX_GeographicBoundingBox. [Libby: a real or imaginary dividing line separating two features]

Resolution: This is the ground area represented by a pixel. For CEOS ARD, this often implies a target spatial resolution (e.g., 10m, 30m) and a consistent grid. [Libby: Smallest change in energy detectable and representable by the number of bits per pixel, where each bit doubles the number of possible values]

A **Pixel** is the fundamental unit of the grid but can be defined in two ways - Pixel as a Point {(pointInPixel = CENTER)}, Pixel as a Cell {cellGeometry} [Libby: Fundamental unit of data collection]

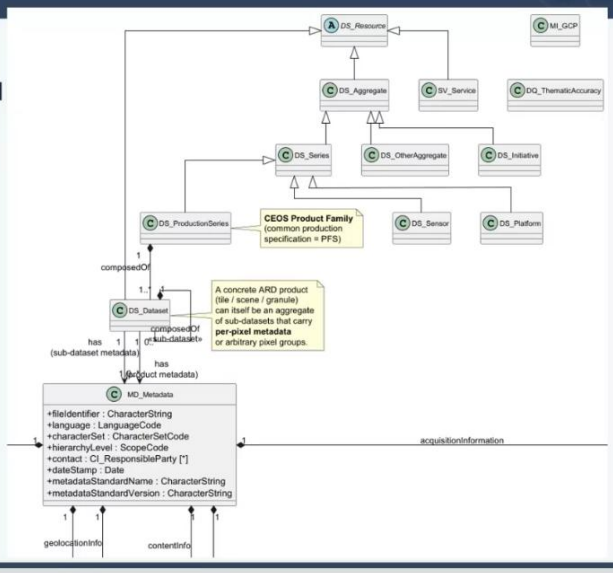
MD_Georectified provides a direct mapping between grid coordinates and geographic coordinates. The centerPoint and cornerPoints are used to describe this mapping, while the pointInPixel attribute explicitly defines the location of the coordinate within the pixel

Every ARD product is anchored by a single metadata root that links to all other components



The Formal Structure: MD_Metadata

Every ARD product is anchored by a single metadata root that links to all other components



LSI-VC-20, 14-16 September 2026 Slide 8

Every ARD product is anchored by a single metadata root that links to all other components

MD_Metadata: The root class.

fileIdentifier, language, characterSet, hierarchyLevel, contact, dateStamp, metadataStandardName, metadataStandardVersion.

Associations:

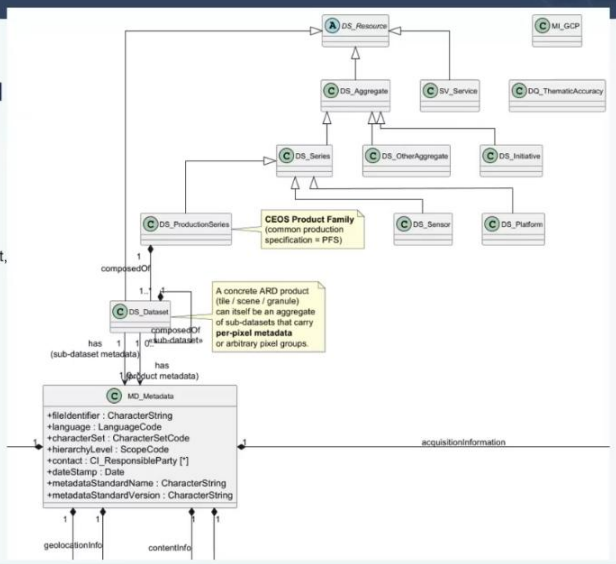
identificationInfo → MD_DataIdentification

spatialRepresentationInfo → MD_SpatialRepresentation

contentInfo → MD_ContentInformation

dataQualityInfo → DQ_DataQuality

acquisitionInformation → MI_AcquisitionInformation

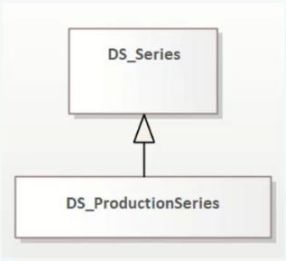


CEOS Product Families are a specialized type of ISO Dataset Series.

Product Family: Collection of datasets that share a common measurement type and adhere to a common production specification (in CEOS-ARD, a Product Family Specification). Corresponds most closely to an ISO 19115-1 DS_ProductionSeries.

Product Family Specification (PFS): A document that defines the Threshold and Target requirements a dataset must meet to belong to a given Product Family and to be recognized as CEOS Analysis Ready Data. It constitutes the “common production specification” as referenced by ISO 19115-1 DS_ProductionSeries.

ISO DS_Series is broader (any common heritage such as sensor, platform, or production specification).CEOS Product Family is narrower and deliberately centred on the production specification (the PFS) and the resulting interoperability for analysis.



The Formal Structure: MD_Metadata

DS_Aggregate: A collection of resources that are managed or grouped together for a common purpose or shared characteristics. It serves as the abstract parent specialization of a resource grouping

DS_Series: A collection of datasets sharing a common heritage, meaning they are related by factors such as originating from the same system, intent, or spatial specification

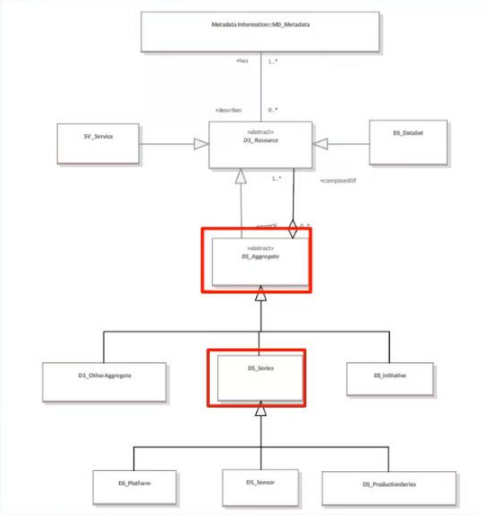
DS_Initiative: A specialized type of resource aggregation grouped around a specific activity, project, mission, or operational campaign

DS_OtherAggregate: A generalized, fallback resource aggregation used when a collection needs to be linked but does not explicitly fit the criteria of a specific activity/initiative or a dataset series

DS_ProductionSeries: A dataset series whose member resources are grouped because they “adhere to a common production specification”

DS_Platform: A dataset series whose component resources are related by the fact that they have been collected or derived from the same physical staging vehicle or carrier (e.g., a specific satellite or aircraft).

DS_Sensor: A dataset series whose component resources are related because they were acquired by the exact same measuring instrument or sensor system.



The Machine-Readable Profile (JSON Schema)

The UML model is translated into a strict JSON Schema for automated exports.

- Required Fields: `ceosard:type`, `ceosard:specification`, `ceosard:specification_version`.
- Conditional Logic (`allOf`):

1. If `ceosard:type` is "optical", then `ceosard:specification` must be one of ["SR", "ST", "AR", "NLSR"].
2. If `ceosard:type` is "radar", then `ceosard:specification` must be one of ["NRB", "POL", "ORB", "GSLC", "INSAR", "CB"].

- STAC Compatibility: `type: [Feature, Collection]`, `stac_version`, `stac_extensions`, `assets`, `links`.

```
"$schema": "https://json-schema.org/draft/2020-12/schema",
"$id": "https://ceos.org/ard/metadata/profile/v1.0/schema.json",
"title": "CEOS-ARD Metadata Profile",
"description": "Minimal formal profile of CEOS-ARD metadata based on ISO 19115-1/19115-2 concepts. Compatible with STAC CEOS-ARD encoding.",
"type": "object",

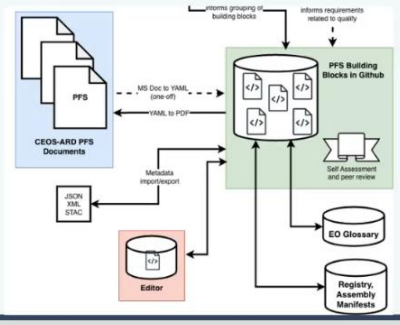
"required": [
  "ceosard:type",
  "ceosard:specification",
  "ceosard:specification_version"
],

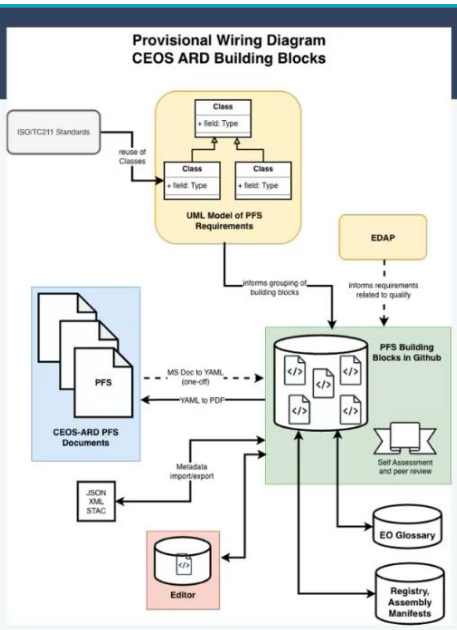
"properties": {
  "ceosard:type": {
    "type": "string",
    "enum": ["optical", "radar"],
    "description": "Sensor type (maps to high-level product family grouping)."
  },
  "ceosard:specification": {
    "type": "string",
    "description": "Product Family Specification identifier (e.g. SR, ST, AR, NLSR, NRB). Corresponds to DS_ProductionSeries."
  },
```

Compliance Checking & Metadata Export

This formalization enables automated validation and standardised export (STAC/JSON/XML).

- Editing/compliance: The pink box in the conceptual diagram - Editor. The JSON schema validates the output of the "PFS Building Blocks in Github".
- Metadata Export: After edits, outputs are standardised JSON, XML, and STAC.
- Assets: The assets object in the JSON schema allows for the "sub-dataset" or "per-pixel metadata" described in the UML note (`DS_Dataset "1" *-- "0..*" DS_Dataset : composedOf «sub-dataset»`).





Final Summary

1. **Standards** (ISO/TC211) provide the vocabulary.
2. **UML Model** formalizes the PFS Requirements.
3. **Building Blocks** (Github) implement the model.
4. **JSON Schema** enforces compliance and helps in editing.
5. **STAC/XML** provides the final, interoperable metadata.

https://github.com/sumitsane/CEOS_ARD_Meta/

