ESA EOPF 101

Table of contents

Welcome

This online book is the go-to resource to learn everything about the EOPF Sample Service by ESA.

To be continued…

Part I

About EOPF

1 Introduction to the Earth Observation Processor Framework

What is the Earth Observation Processor Framework?

The Earth Observation Processor Framework (EOPF) is an initiative led by the European Space Agency (ESA) designed to modernise and harmonise data from the Copernicus Sentinel Missions.

With the upcoming Copernicus Expansion missions in 2028, the amount of data will significantly increase. EOPF is ESA's solution to organise Sentinel data in a way that works seamlessly with modern cloud technology. This will make it easier to find, access, and process the information you need. The new approach provides user-friendly access, simplifies maintenance, and helps keep costs down, guaranteeing reliable access to Sentinel data in the long run.

The Sentinel-1, Sentinel-2, and Sentinel-3 missions are the first to be updated with this new system.

The EOPF Data Model

The EOPF data model has been defined by following a set of principles:

- **Open standards:** Following common and community approved data standards ensure sustainability and user uptake.
- **Interoperability:** Harmonised with a clear and organised structure that describes the data itself.
- **Cloud optimisation:** Designed for efficient access and handling in cloud environments.
- **Conversion flexibility:** Providing tools to adjust the data for different applications.

To achieve this, there are the following key development areas under EOPF:

EOPF product structure

The EOPF prodcut structure defines how information about the data is structured, stored, and connected. It contains the key details we need to know about a dataset, like when it was collected or its level of detail (the attributes). As part of the EOPF, ESA is actively working on a common data structure for Sentinel data products, with the aim to define a common meta-model that can be used nacross all Sentinel and other EO missions. This approach ensures that data from several missions is consistent. Thus, it is easier to combine, relate, and evaluate data from different missions.

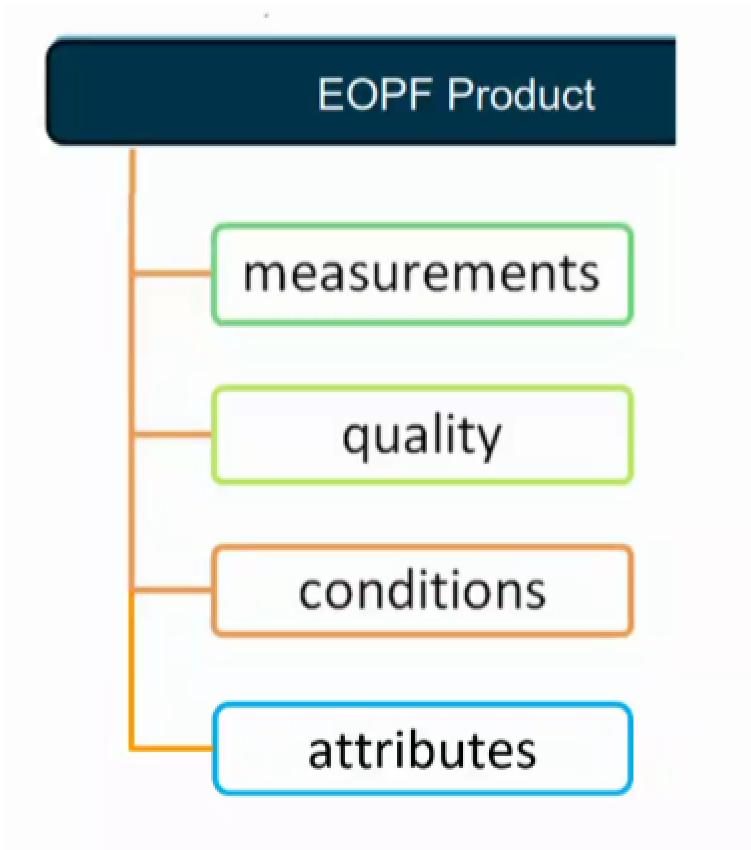

Figure 1.1: EOPF Product structure

The EOPF product structure contains of the following key parts:

- **Measurements:** The actual sensor readings (like how much light is reflected or the temperature), at different levels of detail.
- **Quality indicators:** Details that help understand how reliable the measurements are.

- **Conditions:** Information about the environment or technical aspects when the data was collected.
- **Attributes:** Global metadata, such as when it was acquired and the sensor's orbit.

Metadata Structure

Metadata can be seen as data about data, that provides essential information to understand and use data effectively. Inside the Copernicus ecosystem, it allows us to uniquely identifying each Sentinel product by detailing its technical and geographic coverage information. To make sure the new data system works well and avoids redundancy, the EOPF metadata structure will be organised into:

- Discovery Metadata: Following the metadata structure defined by the SpatioTemporal Asset Catalogue (STAC) which helps to keep things consistent across different missions.
- Processing History Metadata: keeping a record of how the data has been processed.
- Other Metadata: Information like the status of the sensor and details about the satellite's orbit.

> **Note**
>
> *"Learn more about EOPF and STAC here."*

Encoding Structure

An encoding structure can be seen as the specific method used to package and store data and its associated metadata in a digital format. Building on the consistent data structure and clear metadata, the new storage system must be capable of handling various aspects of current Sentinel data (such as manifest files and tile structures from the SAFE format) while remaining fully compatible with cloud environments.

This cloud-friendly architecture will enable instant access to data, efficient processing of massive amounts of data, and seamless integration with other datasets, allowing you to work with data from multiple missions more effectively.

Currently, there are different ways to store geospatial data in the cloud, such as: - Zarr - Cloud-Optimised GeoTIFF (COG) - Hierarchical Data Format (HDF5) with Cloud Adaptations - Apache Parquet (with GeoParquet Extension)

ESA chose `Zarr` as encoding format as it allow for efficient handling of data, like selecting specific areas and combining different datasets easily (for example, by dividing the data into smaller chunks). Learn more about the EOPF Zarr format here.

ℹ Note

Learn more about cloud-optimized geospatial data formats in the Cloud-Optimized Geospatial Data Formats Guide

Processor Development

The way Sentinel data is processed is being updated to take advantage of modern cloud computing. This will make the processing faster and more efficient, while ensuring the scientific quality and accuracy of the Sentinel data remains the same.

ℹ Note

To learn more about the EOPF processor framework, visit https://eopf.copernicus.eu/eopf/

What's next?

In the following chapters, we will introduce the details of the new EOPF data structure, explain the reasons behind ESA's choice of data storage methods, and help the community become comfortable with the new EOPF data format.

2 About Cloud Optimised Formats

Why do we need to cloud-optimise geospatial data formats?

The volume of EO data has grown exponentially in recent years. For ESA, Copernicus alone generates ~16TB daily from the Sentinel missions alone. Traditional file formats, like SAFE (where each file can be hundreds of megabytes), make it slow to get just the specific information needed (for example, data for a single city over a decade). We have to download the entire asset, even if only a small part of it is needed. This might not a problem when you work close to the data.

When moving your Earth Observation processing workflow to the cloud however, your data

Cloud-optimised formats make working with this data much more efficient by allowing:

- **Subsetting extraction**: Avoiding full downloads (optimised Extract-Load-Transform (ELT) operations)
- **Parallel processing**: by working with many parts of the data at the same time.
- **Scalable storage**: Using cloud-native storage systems that can grow as needed.
- **Optimised data latency**: Reducing delays when processing and visualising data.

2.1 What is Data Latency?

Data latency in cloud computing refers to the time it takes for data to be available for processing once requested from cloud storage. For satellite data, this includes the time from when the satellite captures the data to when it is possible to work with it. Reducing latency enables faster access for users without powerful computers, making Sentinel data more accessible to everyone, no matter where they are. It also allows for more detailed analysis, leading to more accurate results because the analysis can use more relevant data.

Minimising latency benefits:

- We can access the data in less time.
- Decisions are made more quickly.
- Combining and analysing different datasets in an efficient way, reducing computing costs.

Modern cloud architectures reduce latency through:

- Storing and processing data in different geographic locations (e.g., AWS regions co-located with ESA ground stations)
- Caching high-demand products, so they are readily available.
- Utilising flexible computing resources that can quickly handle large processing tasks during emergencies

> **Note**
>
> When accessing data over the internet (e.g., cloud storage), latency is high compared to local storage, so it's preferable to fetch lots of data in fewer reads.

2.2 Cloud-Optimised Geospatial Raster Formats

Initiatives like the Cloud Native Geospatial Forum (CNG) help develop ways to use geospatial data efficiently in the cloud, using standard cloud technologies and encouraging open collaboration on data formats. Two of the most important formats for storing and accessing geospatial data efficiently are Cloud-Optimised GeoTIFFs (COGs) and Zarr.

2.2.1 Cloud-optimised GeoTIFF (COG)

COGs are like snapshots of raster data (such as satellite images or elevation maps). This widely used format improves the standard GeoTIFF format by:

- Organising data into **tiles**: Dividing the data into smaller, manageable squares (like 512x512 pixels).
- Including lower-resolution previews: Having pre-generated, less detailed versions of the data.

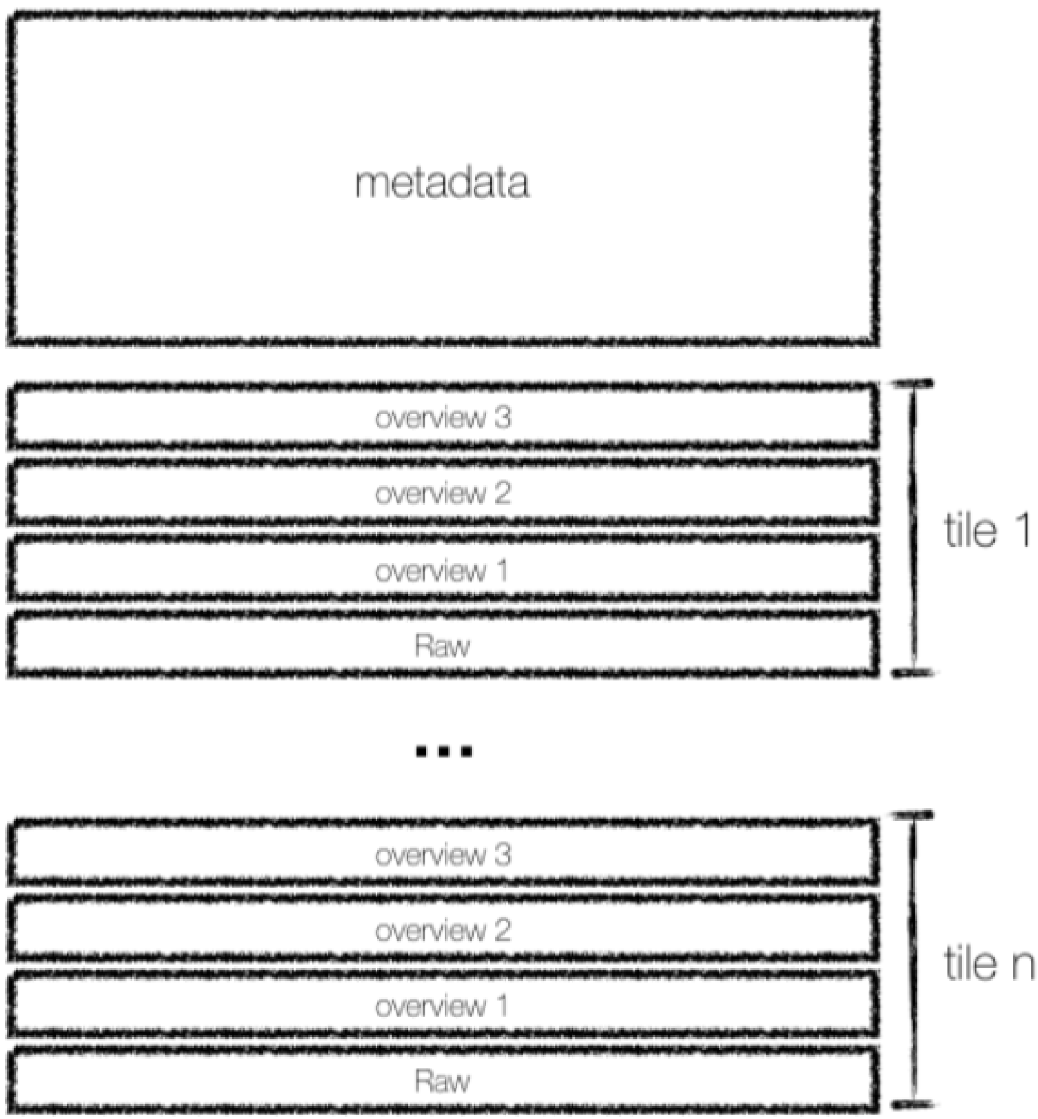

Figure 2.1: COG structure. Retrieved from CNG documentation

A key feature is the **Internal File Directory** (IFD), which acts like an internal index. This allows:

- Retrieving only the parts of the data needed using simple web requests.
- For example, it is possible to access just the tiles covering Paris from a large Sentinel-2 image of Europe.

2.2.2 Zarr

This format is designed for handling large, multi-dimensional datasets (often called "data cubes"). Zarr, developed and maintained by the community, works by:

- Storing data as compressed **chunks** in a flexible way.
- Allowing for efficient indexing and processing of the data in parallel.
- Enabling specific descriptions (metadata) for each part of the data.

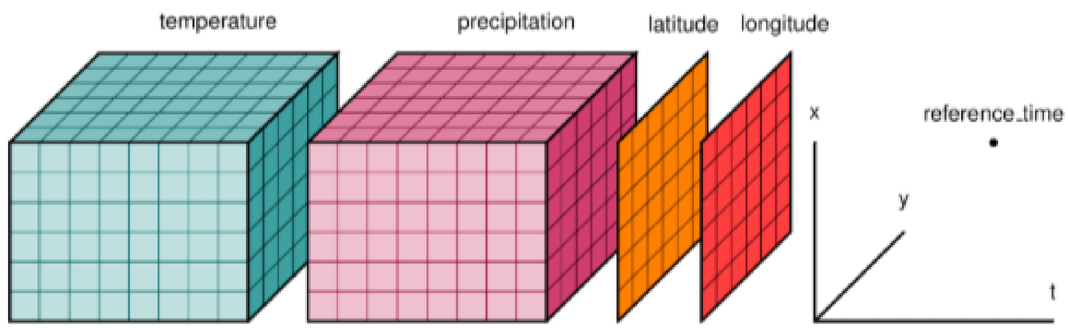

Figure 2.2: Zarr structure. Retrieved from CNG documentation.

For example, Zarr makes it possible to extract temperature data for the summer of 2023 from a climate dataset spanning 50 years without needing to load the entire dataset.

The table below compares some features of COG and Zarr:

Feature	Zarr	COG
Structure	Multi-file chunks	Single file
Access	Parallel	Sequential
Compression	Differently per-chunk	Whole-file
Scales	Multi-scale in single file	Separate, pre-generated lower-resolution files

2.3 When to use COGs versus Zarr?

Based on the structure and capabilities for each format, it is advised to use COGs when:

- Working with raster data (like images or elevation models) that does has sudden changes.
- It is needed to easily visualise or access specific geographic areas without loading the entire dataset.
- Interoperability with existing GIS software is important, as COG is a widely adopted standard.

On the other hand, it is advised to use Zarr when:

- Dealing with large, multi-dimensional datasets that might be updated or modified.
- Performing complex analyses that involve accessing different parts of the data in parallel.
- Efficiently handling different resolutions or variables within a single dataset is required.
- Working in cloud environments that benefit from chunked data storage for parallel processing.

2.4 What's next?

Now that we have an idea of the available cloud-optimised formats and what cloud-optimised means, we will explore the available data sets inside the EOPF available products.

3 The EOPF Available Datasets

The ESA Copernicus Earth Observation Processor Framework uses **Zarr** as the encoding format for the EOPF catalogue. The chapter About Zarr gives you a practical introduction to the sample data being made available as part of the EOPF Sentinel Zarr Sample Service.

The main groups of the Zarr encoding are:

Group	Contents
Measurements	Main retrieved variables
Conditions	Measurement context (geometric angles, meteorological/instrumental data)
Quality	Flags and quality information for measurement filtering
Metadata	Attributes for the compression definition

❗ Important

Note: the re-processing from the legacy SAFE/SEN3 format to Zarr is an ongoing activity as part of the EOPF Sentinel Zarr Sample Service. This page will continuously be updated as soon as new data products are available.

Sentinel-1

Sentinel-1 is radar imaging mission is composed of a constellation of two polar-orbiting satellites providing continuous all-weather, day and night imagery.

3.0.1 Available product Levels

3.0.2 Level 1

These are the first processed versions of the raw data.

Single Look Complex (SLC) SLC images show the Earth's surface as seen by the satellite's radar. Each point has a complex value, giving both the strength and the phase of the signal. The amount of detail depends on how the data was collected.

Ground Range Detected (GRD) GRD products are radar images that have been processed to look like they are viewed from directly above. They have been adjusted for the Earth's shape, and each point represents the signal strength. Some fine details (phase) are lost in this process, but the image has less 'speckling' and a more regular shape.

> **Note**
>
> ESA is currently updating this product. A more detailed structure of this product is available at the EOPF Sentinel-1 Product Documentation

3.1 Sentinel-2

Sentinel-2 acquires optical imagery at high spatial resolution (10m to 60m) over land and coastal waters. The mission supports applications such as agricultural monitoring, emergency management, land cover classifications, and water quality. Sentinel-2 imagery have 12 spectral bands with the following resolutions:

Resolution	Bands
10m	B2, B3, B4, B8
20m	B5, B6, B7, B8a, B11, B12
60m	B1, B9, B10

3.1.1 Available product Levels

3.1.2 Level 2A

Level 2A images are atmospherically corrected Surface Reflectance (SR) images, derived from the associated Level-1C products. These products have been further processed to display surface reflection while adjusting for atmospheric factors (such as air and clouds).

These image products, as well as the Surface Reflectance for the different spectral bands, are resampled at different spatial resolutions. Until now, the ESA, along with the EOPF project, have decided to launch samples for the Level 2A, which encopass the following structure:

Note

ESA is currently updating this product. A more detailed structure of this product is available at the EOPF Sentinel-2 Product Documentation

3.2 Sentinel-3

Sentinel-3 is a mission that regularly measures our Earth's oceans, land, rivers, lakes, ice on land, sea ice, and the atmosphere. Its goal is to keep track of and help us understand how these large parts of our planet change over long periods.

3.2.1 Instrument Suite

Instrument	Name	Measurements	Resolution
OLCI	Ocean and Land Colour Instrumen	t Ocean/land color (21 bands)	300m/1.2km
SLSTR	Sea and Land Surface Temperature Radiometer	Surface temperature (9 channels)	500m/1km
SRAL	Synthetic Aperture Radar Altimeter	Sea surface topography	Variable
SYN	Synergy	Synergistic products from OLCI and SLSTR	N/A

3.2.2 Available product Levels

3.2.3 Level-2

These products process particular information that is helpful for various applications using the Level-1 data. In terms of OLCI, this incorporates factors like the ocean's chlorophyll content and the land surface's reflectivity while accounting for the atmosphere. Sea and land surface temperatures with quality checks are provided by SLSTR. SRAL offers sophisticated ocean data, such as wind speed and wave height. The SYN product analyses the atmosphere by combining information from SLSTR and OLCI.

3.2.3.1 Precise Orbit Data

Information about the satellite's position is available at different times after the data is collected:

- **NRT**: Near Real-Time (GPS-based)
- **STC**: Short Time-Critical (1-day latency)
- **NTC**: Non-Time-Critical (post-processed)

> **Note**
>
> ESA is currently updating this product. The product structure is described in more detail at the EOPF Sentinel-3 Product Documentation

In the following chapter, the `zarr` specific structure for each of the products will be explored, allowing us to understand better, how to integrate the new format into Earth Observation workflows.

Glossary

References