
Cloud Optimized Formats

2.9.2024



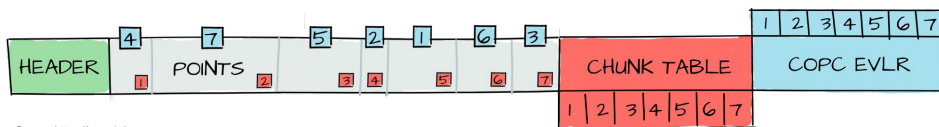
Agenda

- Context / Cloud Optimized Formats
- Cloud Optimized Formats - State
- Outlook

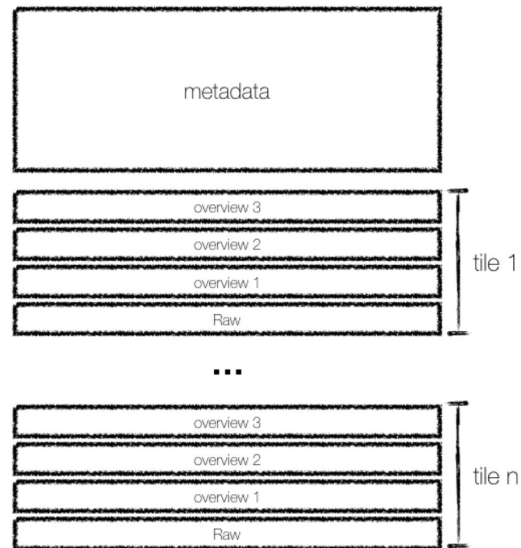


Context / Cloud Optimized Formats

- Little Server Infrastructure (Serverless)
- Streaming / RANGE Requests
- Smart Metadata / Data Organisation



Source: <https://copc.io/>

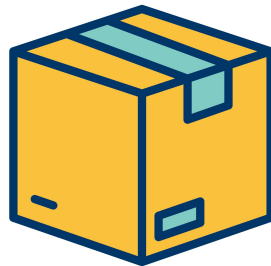


Source: <https://guide.cloudnativegeo.org/cloud-optimized-geotiffs/intro.html>



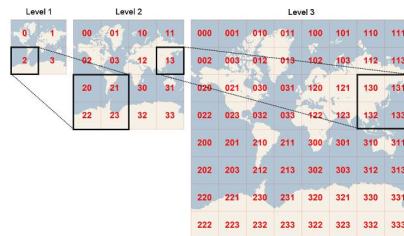
Context / Cloud Optimized Formats

- Optimized for publishing static data
- Coarse access control (on dataset level)
- Optimized for big data, for smaller data readable formats like CSV or JSON can be considered
- Integrates well with CDN, can well handle sudden peaks



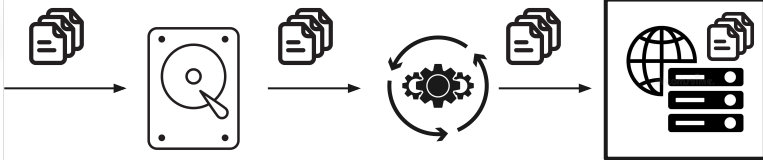
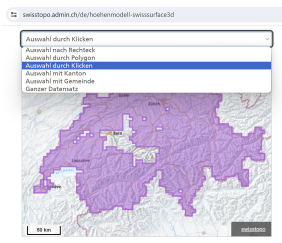
Source: <https://www.vecteezy.com>

VS

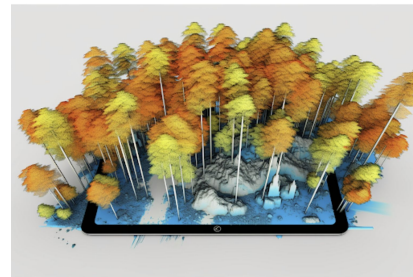
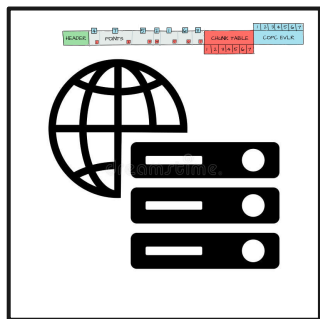


Source: <https://developer.myptv.com/en/documentation/raster-maps-api/concepts/tiles>

Context / Cloud Optimized Formats



VS



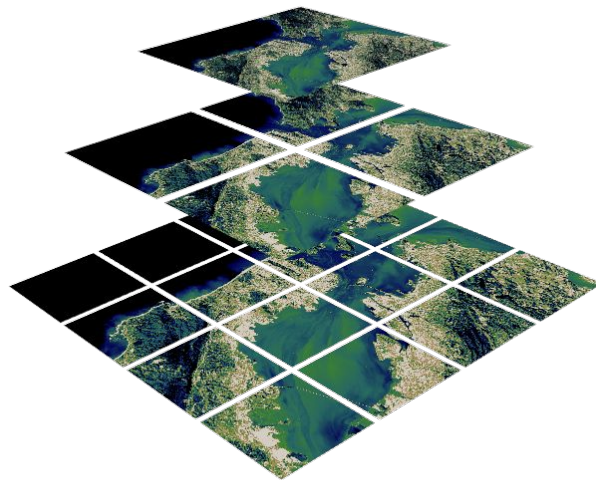
Cloud Optimized Formats State



Raster Data - COG

- **Cloud Optimized GeoTiff**
- A natural candidate as tiling and overviews can be generated
- Backwards compatible
- Can be used even for non-cloud use cases
- OGC Standard

Recommended



Source: <https://www.kitware.com/deciphering-cloud-optimized-geotiffs/>

Vector Data - Challenges

1

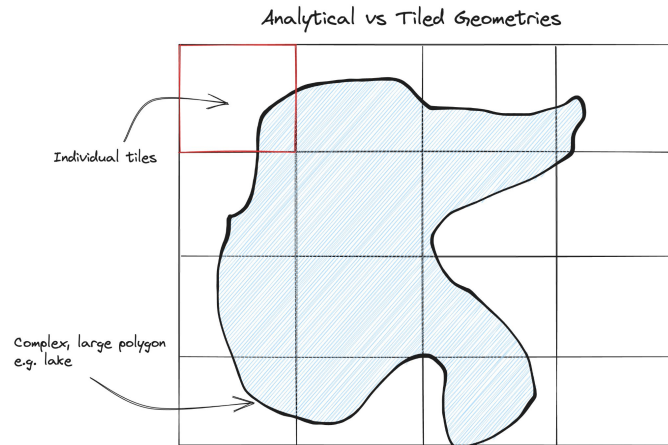
Generalization

Which objects to eliminate?
Which vertices to eliminate? Topology preservation etc?
Which attributes to eliminate?

2

Tiling

Large objects can span over multiple tiles

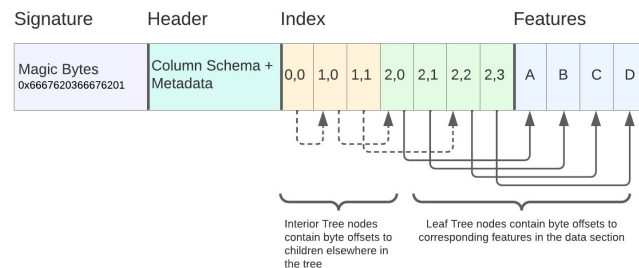




Vector Data - FlatgeoBuf

- Stores geometries (OGC simple features) with attributes
- Good support in open source geospatial tools (GDAL, Javascript), no native ArcGIS support
- Is uncompressed: transfer-encoding or SOZip
- OGC Community standard candidate

Contender





Vector Data - GeoParquet

- Stores geometries (OGC simple features) with attributes
- Column based, optimized for specific access patterns, column metadata available
- Availability in tools not yet perfect (GDAL extra driver)
- Needs a lot of expertise to not shoot your foot
- OGC standard candidate

Breaking news: version 1.1 has now spatial indexes

	Column 1	Column 2	Column 3	Column 4	Column 5
	Product	Customer	Country	Date	Sales Amount
Row Group 1	Ball	John Doe	USA	2023-01-01	100
	T-Shirt	John Doe	USA	2023-01-02	200
Row Group 2	Socks	Maria Adams	UK	2023-01-01	300
	Socks	Antonio Grant	USA	2023-01-03	100
Row Group 3	T-Shirt	Maria Adams	UK	2023-01-02	500
	Socks	John Doe	USA	2023-01-05	200

Contender



Vector Data - PMTiles

- Optimized for visualization
- Is a generic container
- Supports tiling and overviews (through MVT)
- Also supports raster data, used less common
- No (OGC) standard

Contender





Vector Data - COSh

- Cloud Optimized Shapefile
- Based on principles of spinning harddrives, where seeks were expensive and loading data in chunks was cheap
- Backwards compatible
- Plenty of problems we don't want to deal with any more

Not Recommended

I 
SHP



Multidimensional Data - GeoZARR

- Optimized for N-dimensional data, hyperspectral scanners, temporal
- Used in meteorology, earth observation
- Based on Zarr (which is an OGC standard without spatial properties)
- Discussed in an OGC working group

Contender





Point Cloud Data - COPC

- Cloud Optimized Point Cloud
- Point Cloud Data is – like raster data – a natural fit for hierarchic structures
- Based on the LAZ 1.4 specification (backwards compatible)
- Support in tools getting better and better

Contender

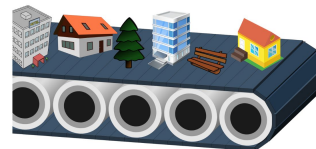
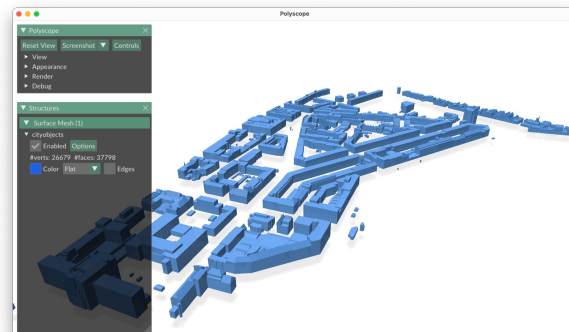


COPC.io
Cloud Optimized Point Cloud

CityJSONSeq (CityJSON Text Sequences)

- Cloud Optimized CityJSON (3D Building models)
- JSON organized/optimized for streaming
- Based on CityJSON 2.0 (OGC Standard)
- Support in tools getting better and better

Contender





This and that ...

- SOZip for storing compressed data in a seek optimized way
- 3D Data: still on the way, 2 standards, both not streaming optimized (for what I am aware), some efforts are underway
<https://github.com/CesiumGS/3d-tiles/issues/399>
- Catalogues: STAC and OGC API Record are supporting the cloud optimized ecosystem
<https://github.com/stac-utils/stac-crosswalks/>



Summary

Standards

OGC is currently actively working on standardizing formats. It can be expected that in the next months many standards will be formalized.

Support

The ecosystem of tools is growing and expanding support. Open source tools faster in embracing support.

Ecosystem

Data is becoming more and more available. The ecosystem grows in an agile way,

Questions?

