EURO DATA CUBE

Euro Data Cube and Zarr

Brockmann Consult / Euro Data Cube

norman.fomferra@brockmann-consult.de

SINERGISE BROCKMANN CONSULT GMBH EOX gisat planet. SENTINEL Hub

What EDC offers and requires

EURO DATA CUBE

- EDC offers to users **data cubes** and **data cube services**
- EDC data cubes are **analysis-ready**, **cloud-ready**, **multi-variate**, **multi-dimensional**, **gridded** datasets
- EDC converts existing data sources (EO, model, GIS, ... data) into
  - … volatile data cube views of existing data web APIs, e.g. Sentinel HUB;
  - … static data cube instances in cloud object storage, e.g. cube subsets by users.
- EDC data cube datasets
  - … may be of any size in any dimension, still allow users to efficiently access subsets of data;
  - … must use a comprehensive and understandable data and metadata model that conforms to existing community de-facto standards;
  - … must be directly accessible by clients using solely open source format specifications and software packages.

EDC uses Zarr, for many reasons!

EURO DATA CUBE

- The Zarr format is open, easy to understand, and easy to implement.
- Zarr allows data arrays to be chunked, for efficient access to subsets of ultra large datasets. Chunks are compressed. No-data chunks may be omitted entirely. Therefore, accessing dataset subsets is fast and efficient.
- Zarr has been designed with cloud readiness in mind. Ideal for cloud object storage.
- Zarr Python implementation (zarr.readthedocs.io)
  - opens datasets from simple mutable key-value dictionaries. This allows for numerous persistent storage backends (S3, Redis, MongoDB, DBM). And this allows for creating data cube views from existing data APIs.
  - utilises chunked, lazy-loaded dask arrays and are therefore numpy-compatible.
  - is utilized by popular xarray package. EDC uses xarray datasets to represent data cubes in Jupyter notebooks and user code.
- Zarr is now available for a growing number of programming languages. Brockmann Consult is actively developing Zarr for Java (jzarr.readthedocs.io).
- Zarr has been proposed as new OGC community standard.
- Zarr has an active user and developer community.
- Zarr can be easily extended to represent multi-scale datasets (i.e. image pyramids), such that online mapping clients can visualise and map large datasets directly from cloud sources. Brockmann Consult is actively developing such clients, e.g. the xcube Viewer and the ESA CCI Toolbox Cate.