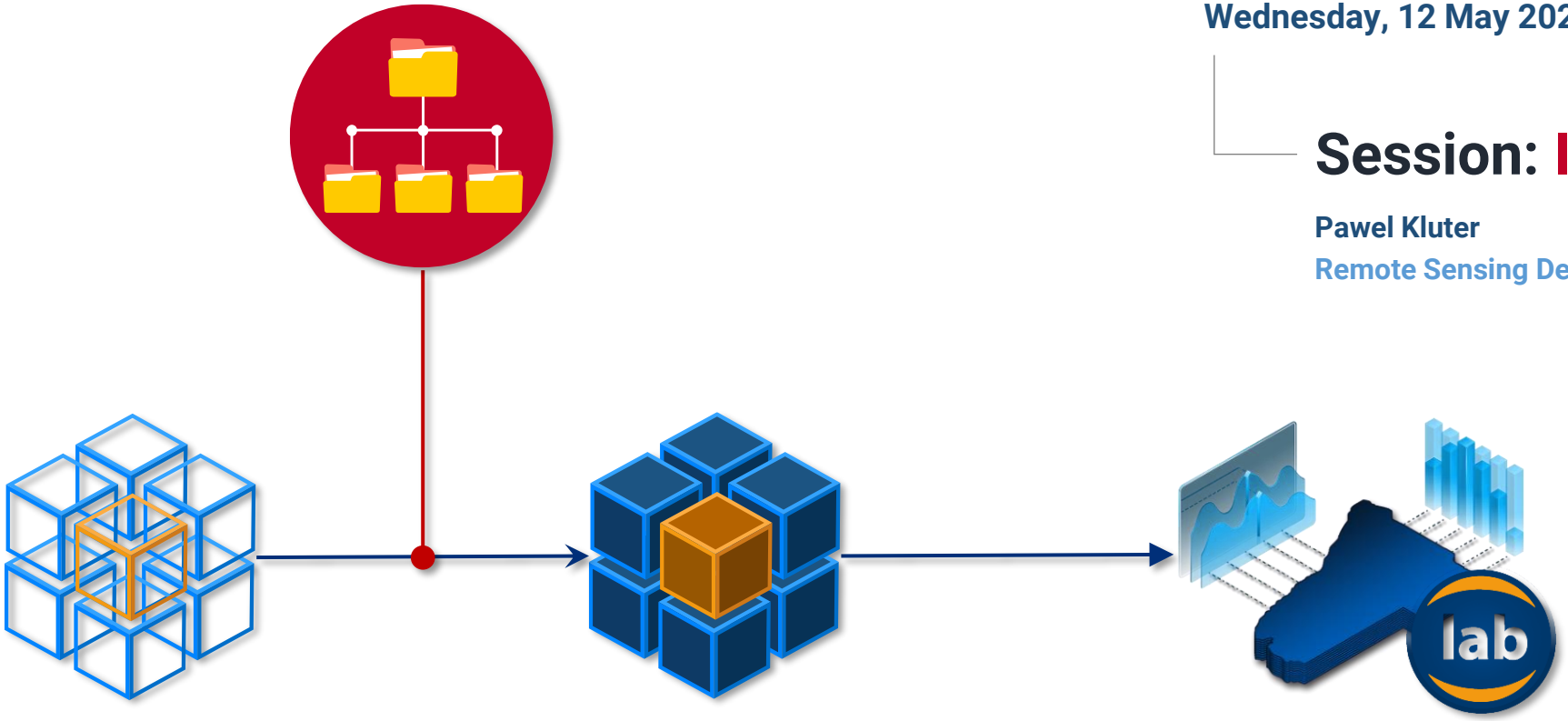
SDGs4Namibia DC Training

Wednesday, 12 May 2021 **2 – 4 pm CET**

Session: Ingestion



Pawel Kluter
Remote Sensing Department, JMU Würzburg

giz

Namibia Statistics Agency

Ingestion FAQs

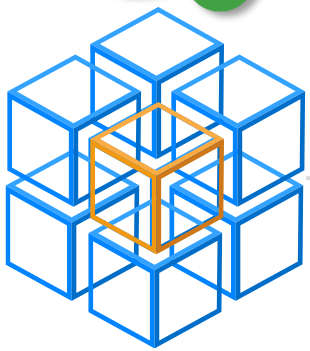
QUESTION:

Does ingestion handle preprocessing or does data need to be processed before ingestion?

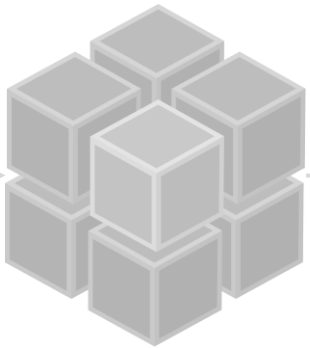
ANSWER:

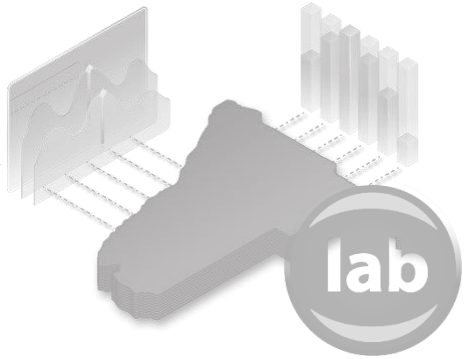
The ingestion process is simply a reprojection and resampling process for existing data. Data should be preprocessed before ingestion.

www.opendatacube.org/faq



Install DC Packages
&
Dependencies

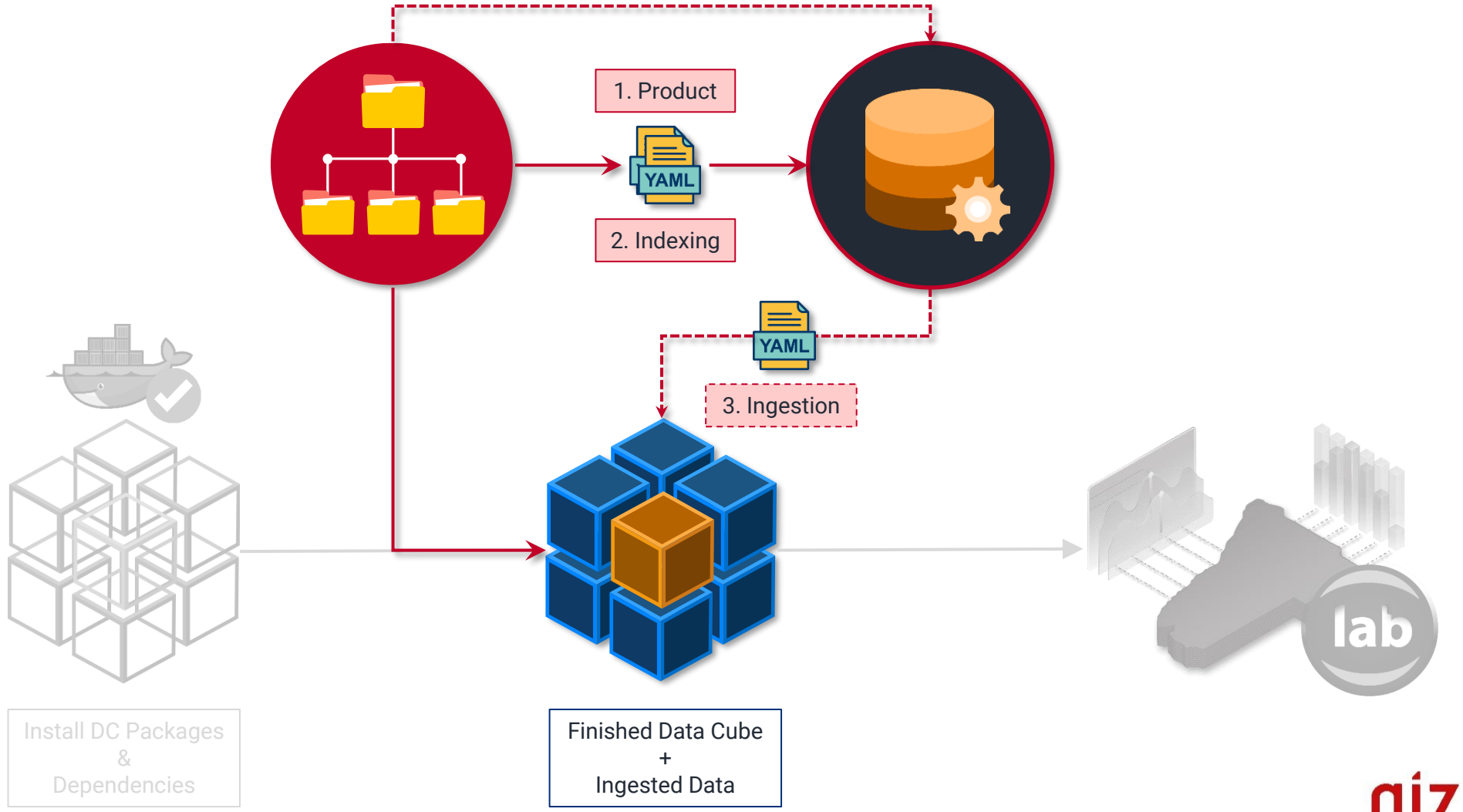




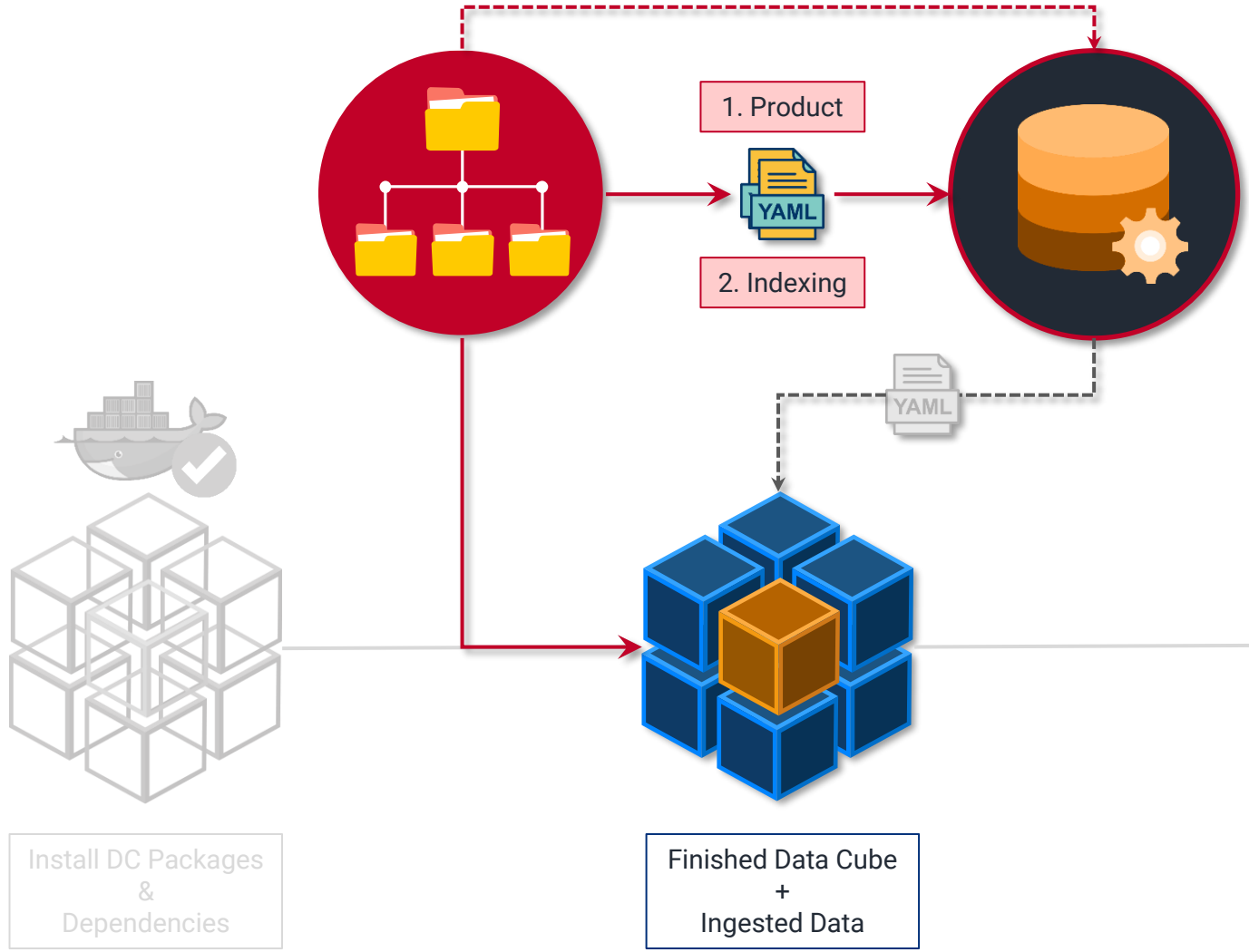
giz

Namibia Statistics Agency

Core Ingestion Concept



Core Ingestion Concept



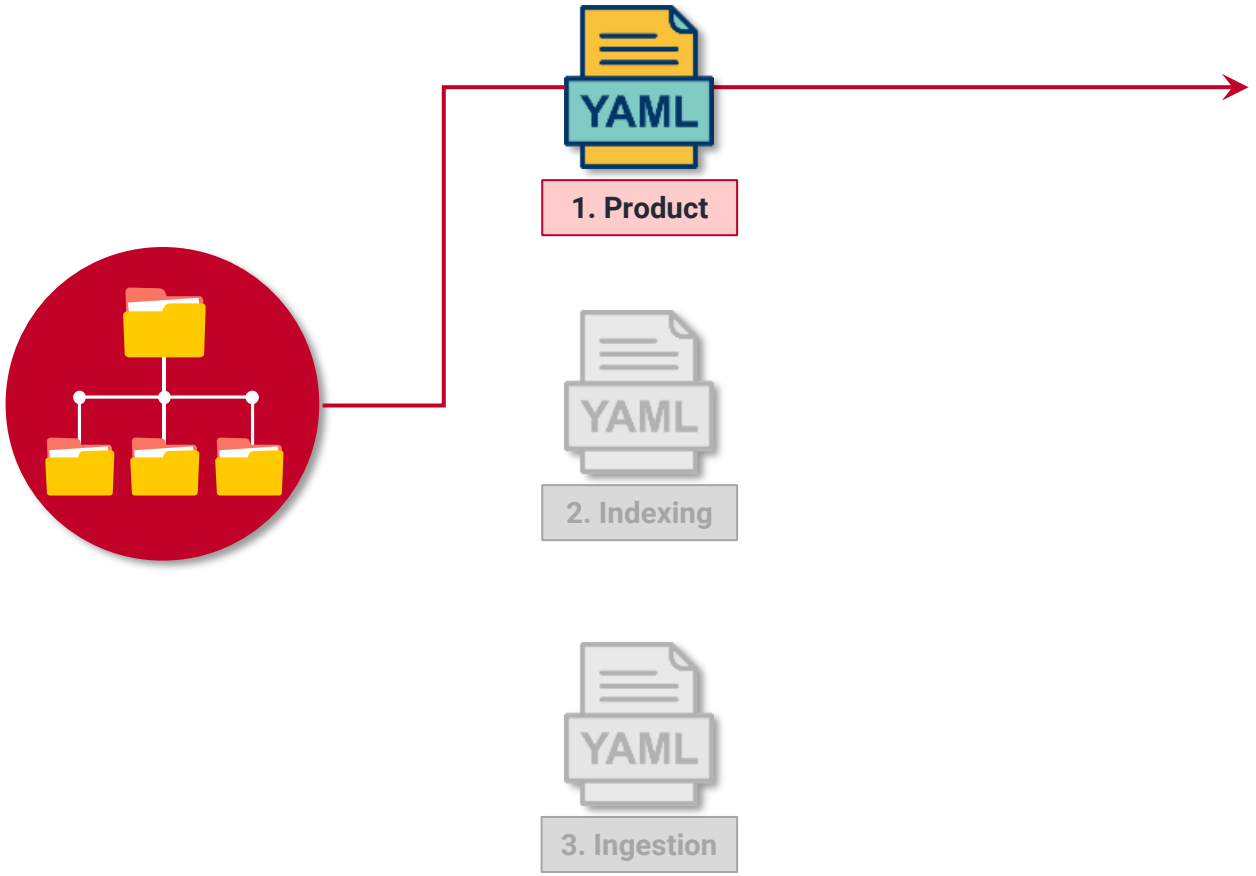
no Ingestion | ARD

ARD – Analysis Ready Data

- **Atmospheric Correction**
- **Pixel Alignment**
- **Sensor Alignment**
- **Usable Data Masks**
- **Database compatible**

Ingestion Product YAML

Dataset Preparation: MODIS – Global Waterpack (DLR)



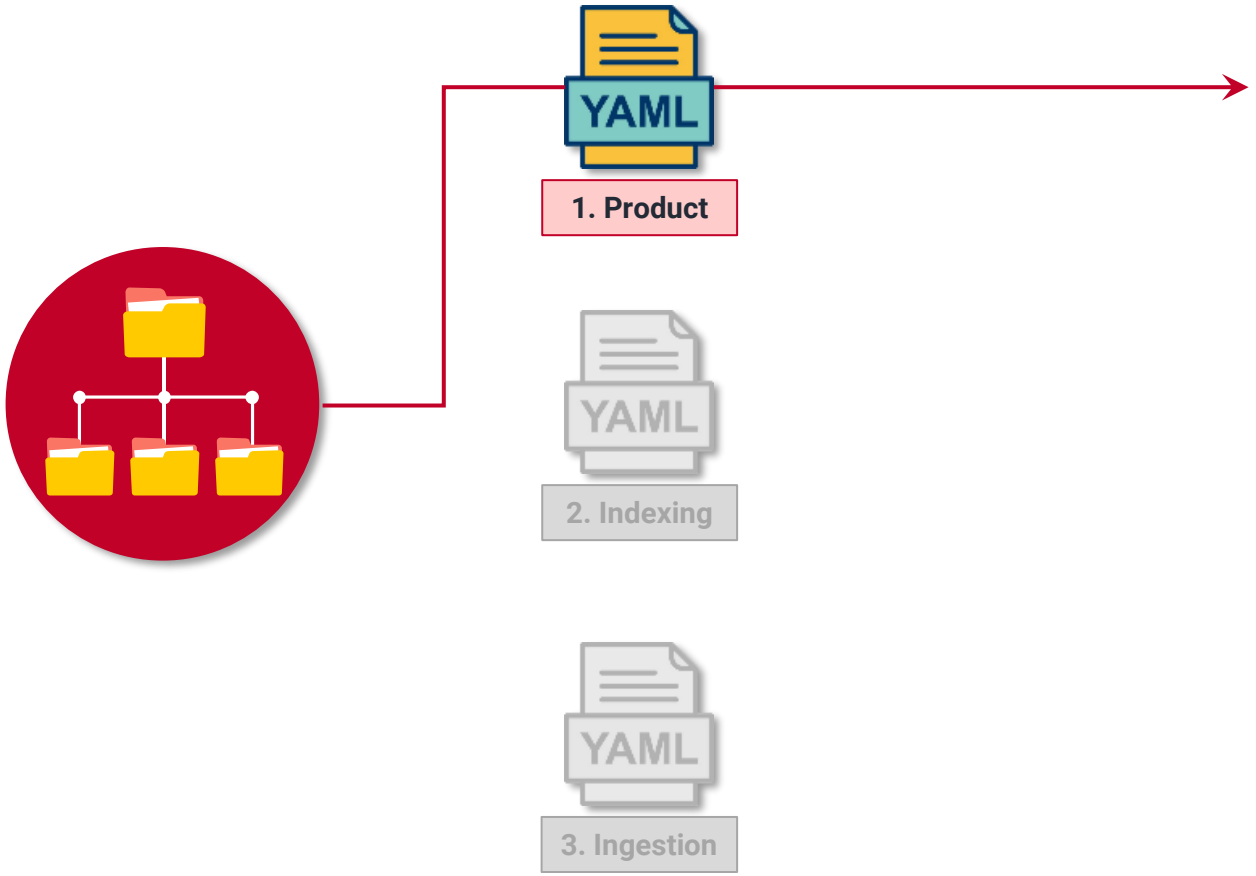
25 lines (23 sloc) | 446 Bytes

```yaml
name: global_waterpack
description: daily inland surface water extent
metadata_type: eo

metadata:
  platform:
    code: MODIS
  instrument:
    name: AQUA_TERRA
  product_type: GWP
  format:
    name: GeoTIFF

measurements:
    - name: 'GWP'
      dtype: uint8
      nodata: 255
      units: '1'
      flags_definition:
        scl:
          bits: [0,1]
          description: GWP
          values:
            0: no water
            1: water
```

eo2cube repository: github.com/eo2cube/eo2cube_config

Ingestion Product YAML

Dataset Preparation: MODIS – Global Waterpack (DLR)



25 lines (23 sloc) | 446 Bytes

```yaml
name: global_waterpack
description: daily inland surface water extent
metadata_type: eo

metadata:
  platform:
    code: MODIS
  instrument:
    name: AQUA_TERRA
  product_type: GWP
  format:
    name: GeoTIFF

measurements:
    - name: 'GWP'
      dtype: uint8
      nodata: 255
      units: '1'
      flags_definition:
        scl:
          bits: [0,1]
          description: GWP
          values:
            0: no water
            1: water
```

xarray.Dataset

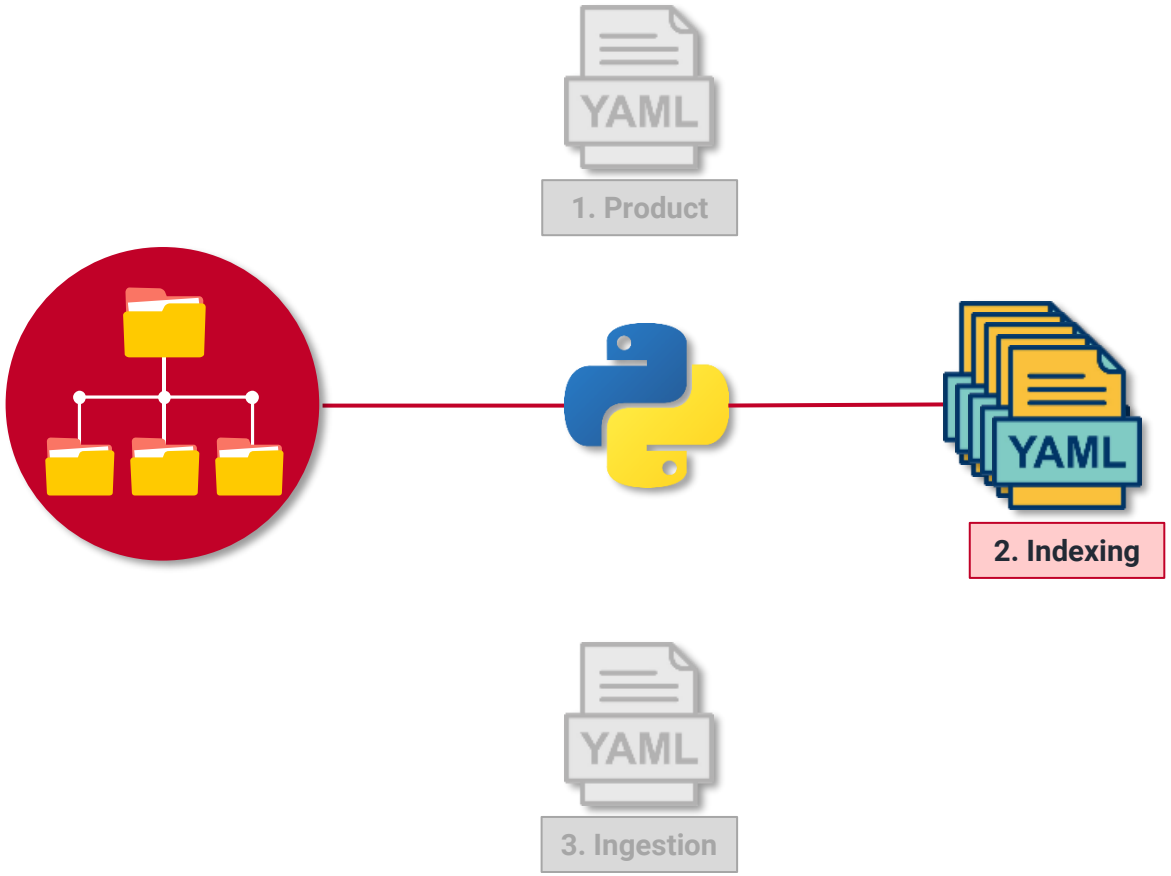
xarray.Dataset

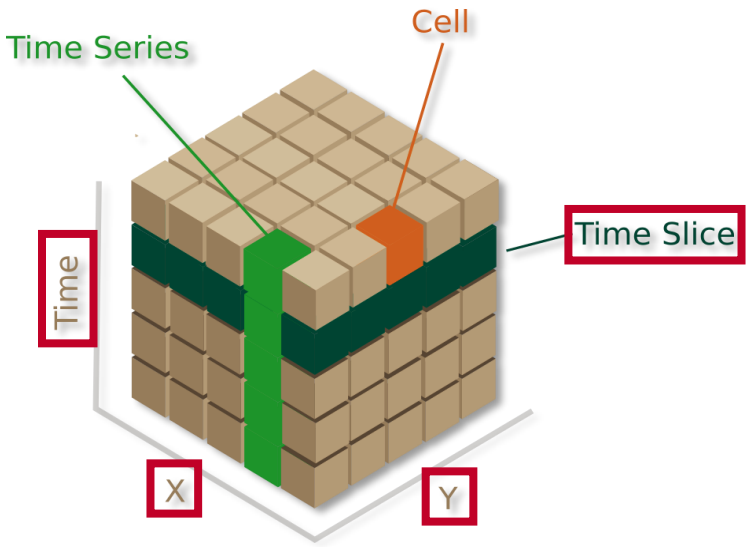
xarray.Dataset

eo2cube repository: github.com/eo2cube/eo2cube_config

Ingestion Index YAMLs

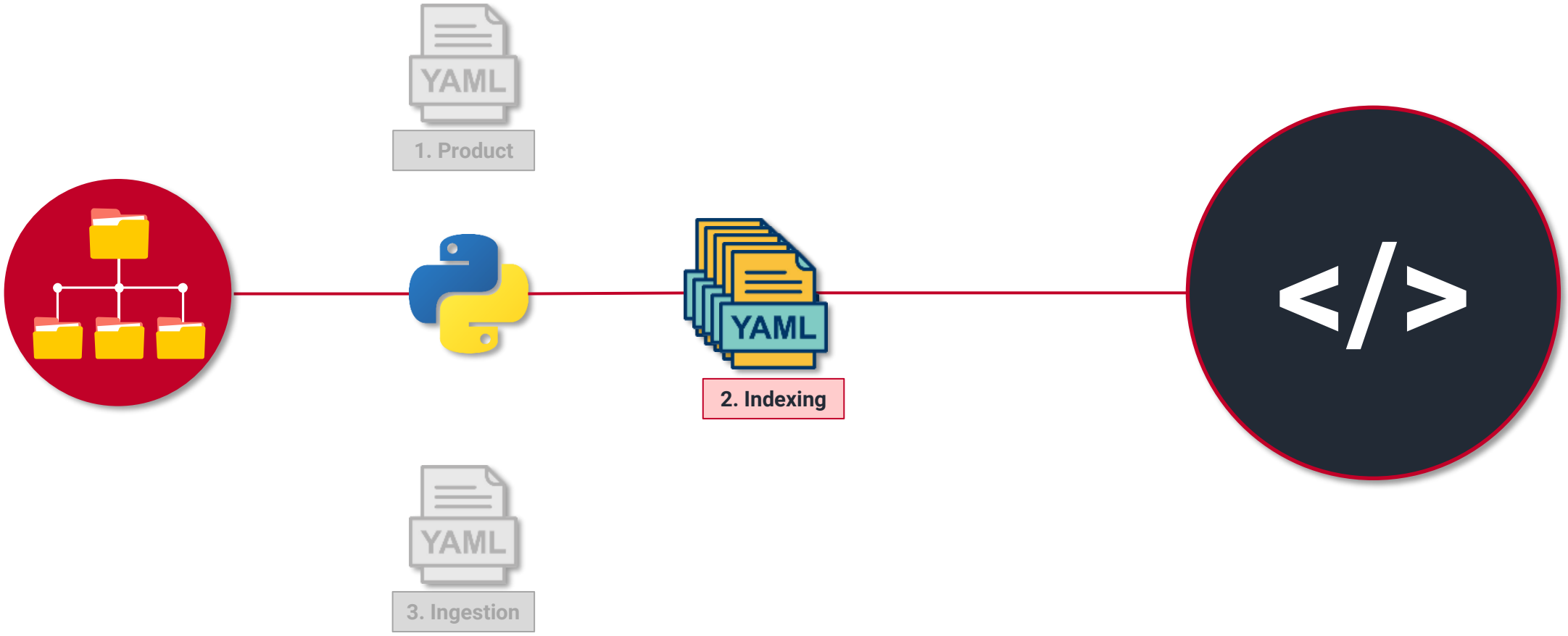
Dataset Preparation: MODIS – Global Waterpack (DLR)





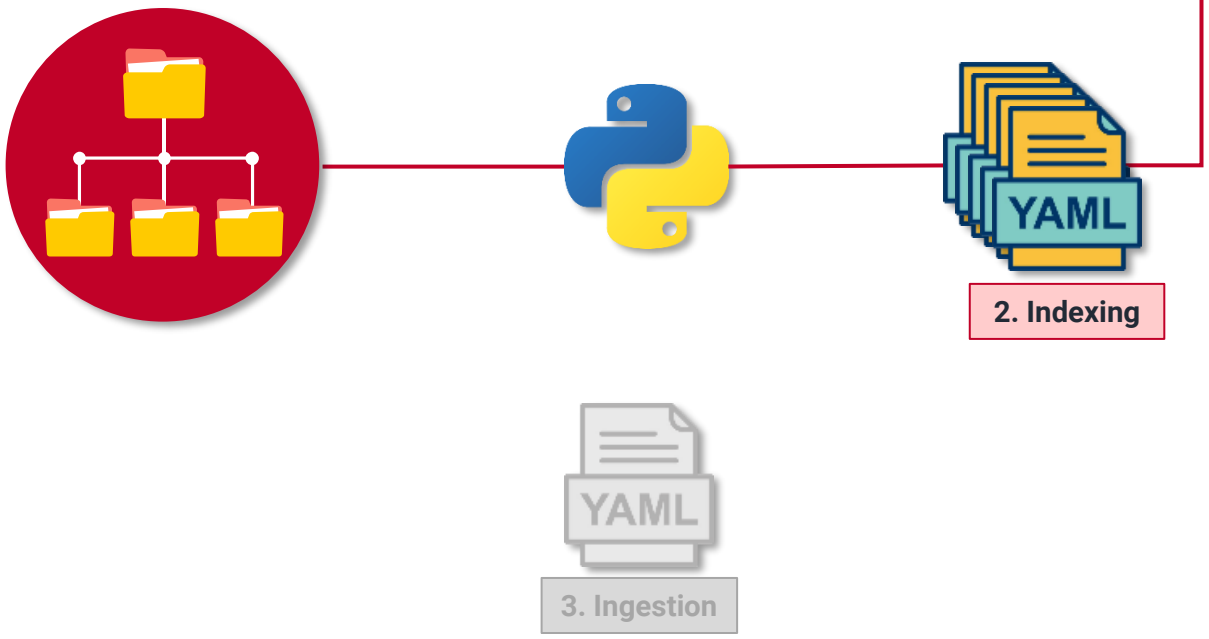
Ingestion Index YAMLs

Dataset Preparation: MODIS – Global Waterpack (DLR)



Ingestion Index YAMLs

Dataset Preparation: MODIS – Global Waterpack (DLR)



53 lines (53 sloc) | 1.6 KB

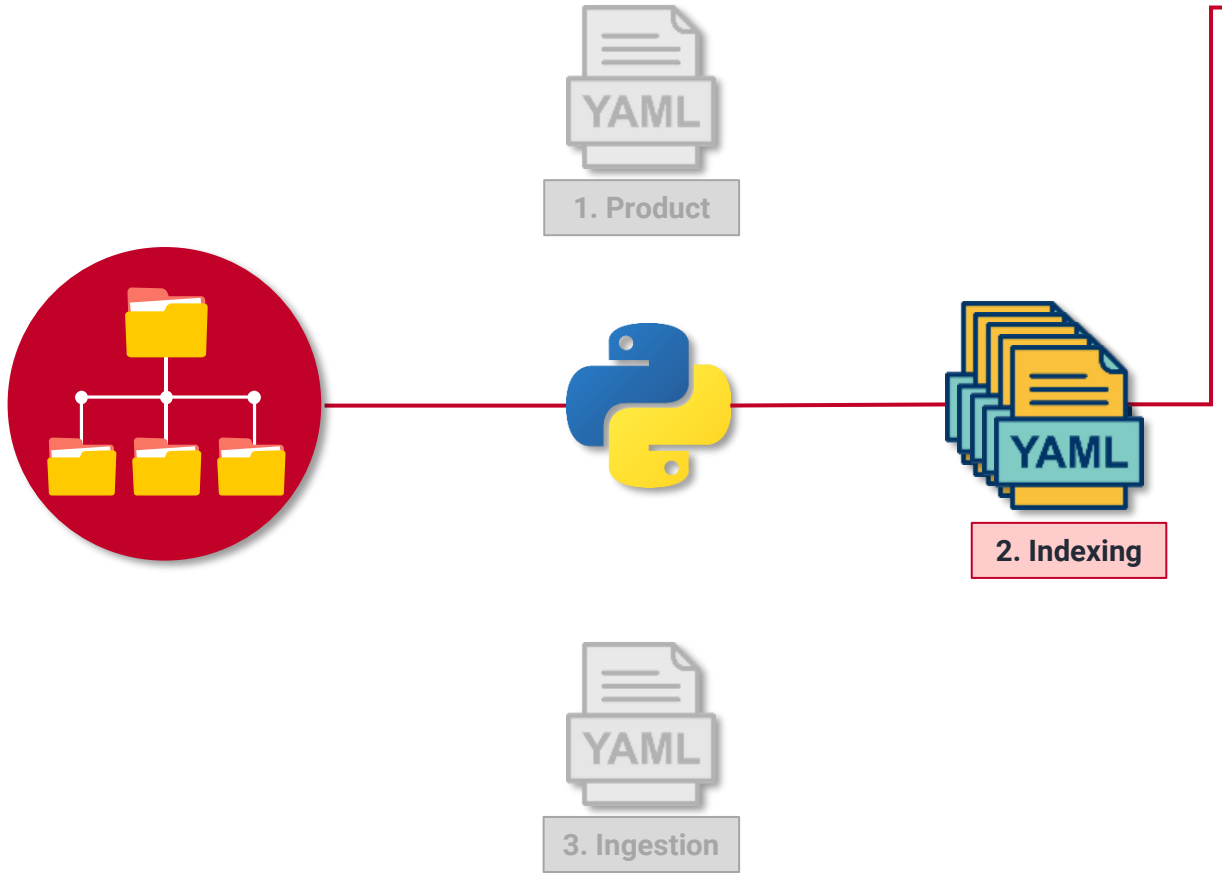
```yaml
acquisition:
  groundstation:
    code: MODIS
creation_dt: '2013-01-02T00:00:00.000Z'
extend:
  center_dt: '2013-01-02T00:00:00.000Z'
  coord:
    ll:
      lat: 10.2644892
      lon: -30.1646205
    lr:
      lat: 23.0943147
      lon: -29.9999467
    ul:
      lat: 10.6432518
      lon: -19.9464572
    ur:
      lat: 22.4519103
      lon: -19.8433909
  from_dt: '2013-01-02T00:00:00.000Z'
  to_dt: '2013-01-02T00:00:00.000Z'
format:
  name: GeoTIFF
```

time relation

eo2cube repository: github.com/eo2cube/eo2cube_config

Ingestion Index YAMLs

Dataset Preparation: MODIS – Global Waterpack (DLR)



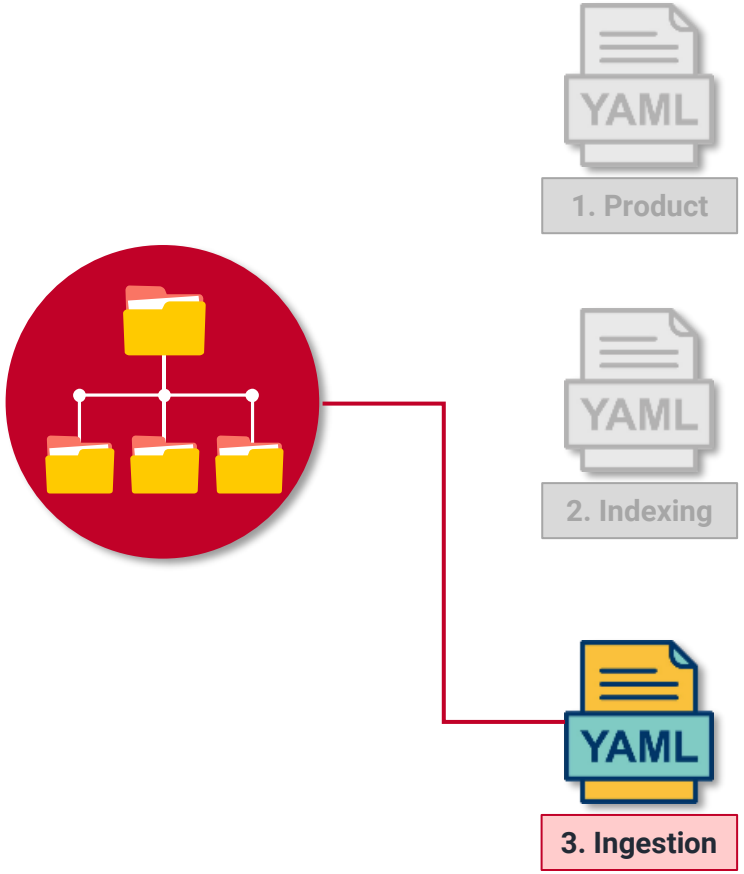

```yaml
grid_spatial:
  projection:
    geo_ref_points:
      ll:
        x: 43767.818
        y: 6653485.712
      lr:
        x: 1281985.198
        y: 6653485.712
      ul:
        x: 43767.818
        y: 7788518.31
      ur:
        x: 1281985.198
        y: 7788518.31
    spatial_reference: PROJCS["WGS 84 / UTM zone 33S",GEOGCS["WG
id: 5f48decf-19bc-4906-b3cc-6f1b0a32f630
image:
  bands:
    GWP:
      layer: 1
      path: /datacube/GlobalWaterpack/2013/reproj_INT09GQ.A20130
instrument:
  name: AQUA_TERRA
lineage:
  source_datasets: {}
platform:
  code: MODIS
processing_level: L2
product_type: GWP
```

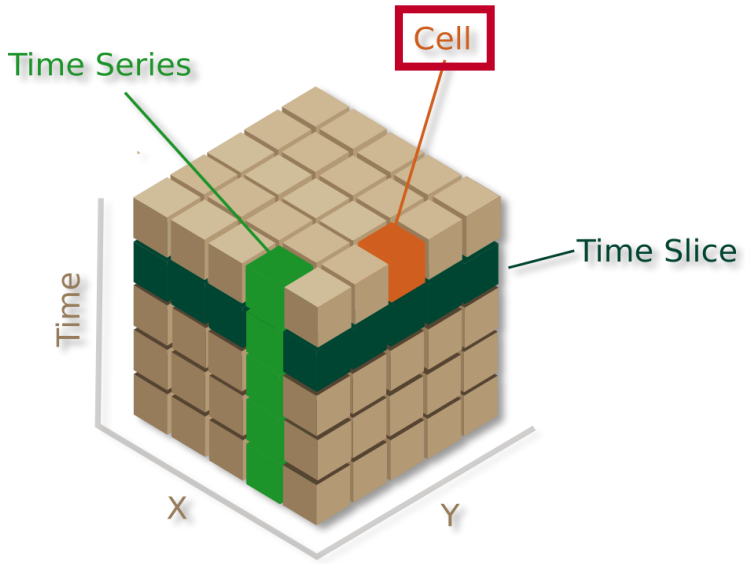
x & y extent

xarray.Dataset

eo2cube repository: github.com/eo2cube/eo2cube_config

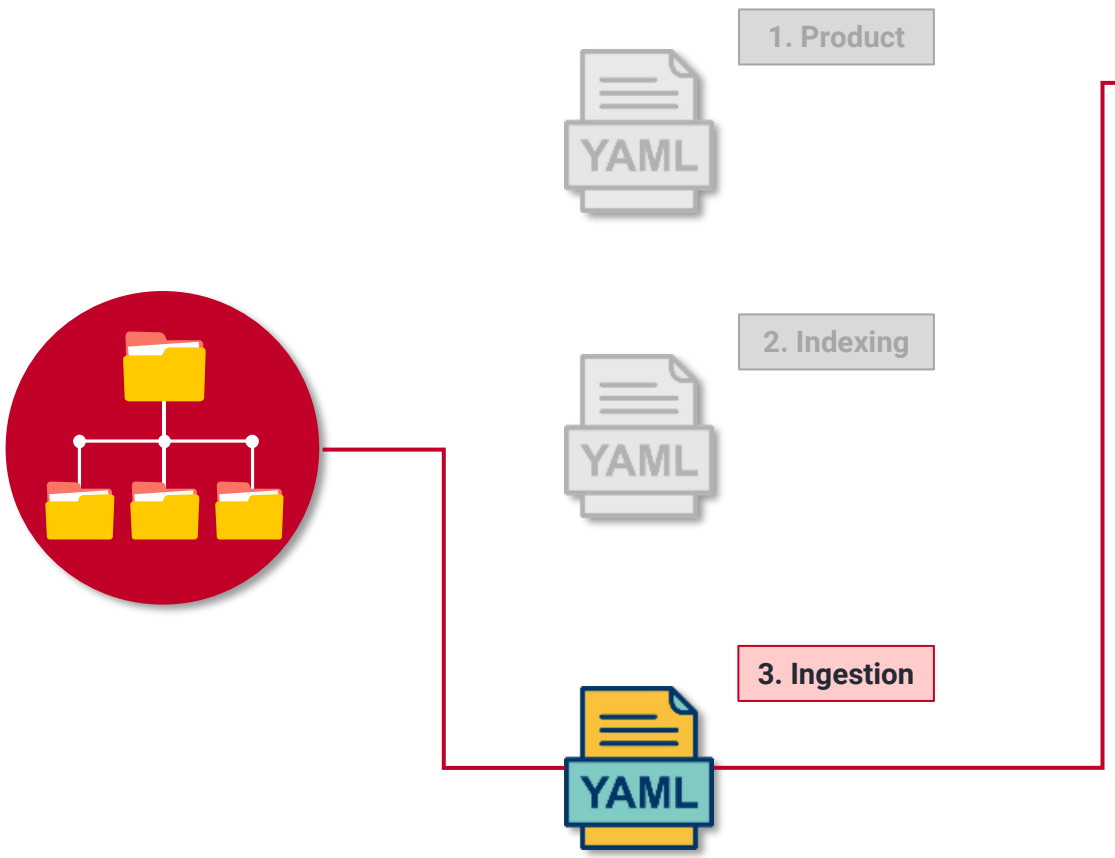
Dataset Preparation: MODIS – Global Waterpack (DLR)





Ingestion Ingest YAML

Dataset Preparation: MODIS – Global Waterpack (DLR)



52 lines (45 sloc) | 1.19 KB

```yaml
source_type: global_waterpack
output_type: modis_global_waterpack

description: daily inland surface water extent, EPSG:32733

location: '/datacube/ingested_data/'
file_path_template: 'modis_global_waterpack/{tile_index[0]}/{tile_index[1]}
global_attributes:
  title: Namibian Data Cube
  summary: Global Waterpack
  source: German Aerospace Center (DLR)
  institution: JMU Wuerzburg
  instrument: AQUA_TERRA
  cdm_data_type: Grid
  keywords: NAMIBIA,GWP
  platform: MODIS
  processing_level: L2
  product_version: '1.0.0'
  product_suite: Global Waterpack
  project: Namibian Data Cube
  coverage_content_type: classification
  references: doi(10.1016/j.rse.2017.06.045)
  license: doi(10.1016/j.rse.2017.06.045)
  naming_authority: German Aerospace Center (DLR)
  acknowledgment: see license
```

eo2cube repository: github.com/eo2cube/eo2cube_config

Ingestion Ingest YAML

Dataset Preparation: MODIS – Global Waterpack (DLR)

YAML
1. Product

YAML
2. Indexing

YAML
3. Ingestion

```
storage:
  driver: NetCDF CF

  crs: EPSG:32733
  tile_size:
          x: 500000.0
          y: 500000.0
  resolution:
          x: 250
          y: -250
  chunking:
      x: 200
      y: 200
      time: 1
  dimension_order: ['time', 'y', 'x']

fuse_data: copy

measurements:
    - name: GWP
      dtype: uint8
      nodata: 255
      resampling_method: nearest
      src_varname: 'GWP'
      zlib: True
```

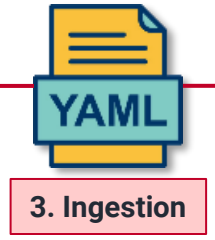
reprojection

tile size: 500 x 500 km

overlaping data fusion

xarray.Dataset

Dataset Preparation: MODIS – Global Waterpack (DLR)




```
storage:
  driver: NetCDF CF

  crs: EPSG:32733
  tile_size:
          x: 500000.0
          y: 500000.0

fu

me

      dtype: uint8
      nodata: 255
      resampling_method: nearest
      src_varname: 'GWP'
      zlib: True
```

reprojection

tile size: 500 x 500 km

xarray.Dataset

Dataset Preparation: MODIS – Global Waterpack (DLR)

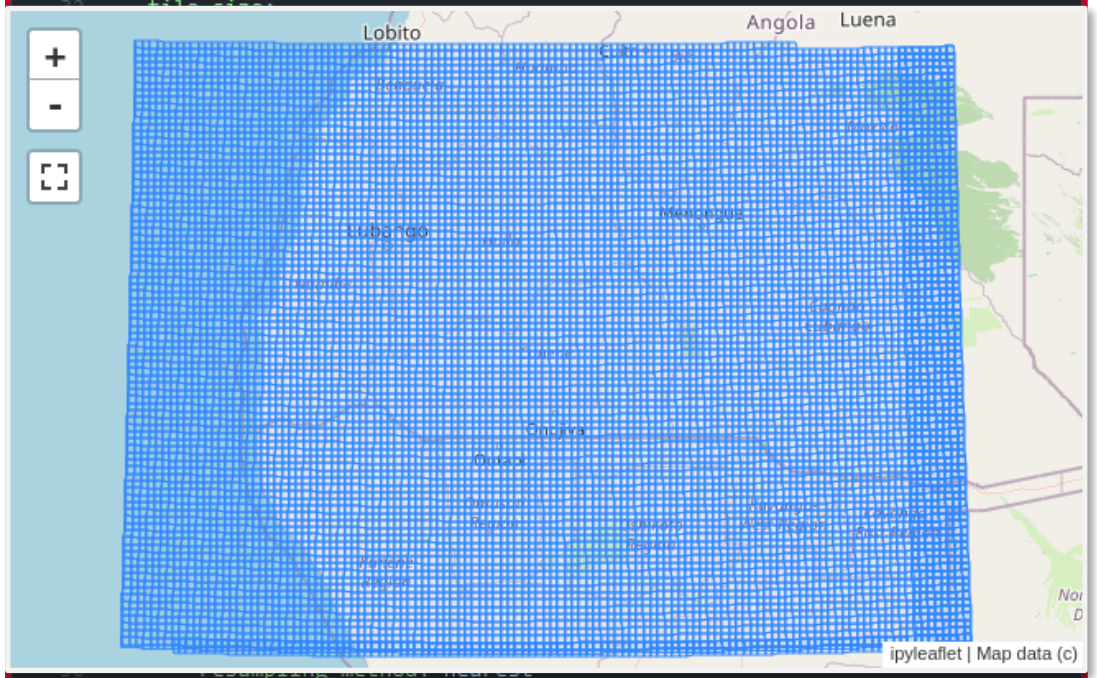
YAML
1. Product

YAML
2. Indexing

YAML
3. Ingestion

```
28  storage:
29    driver: NetCDF CF
30
31    crs: EPSG:32733
```

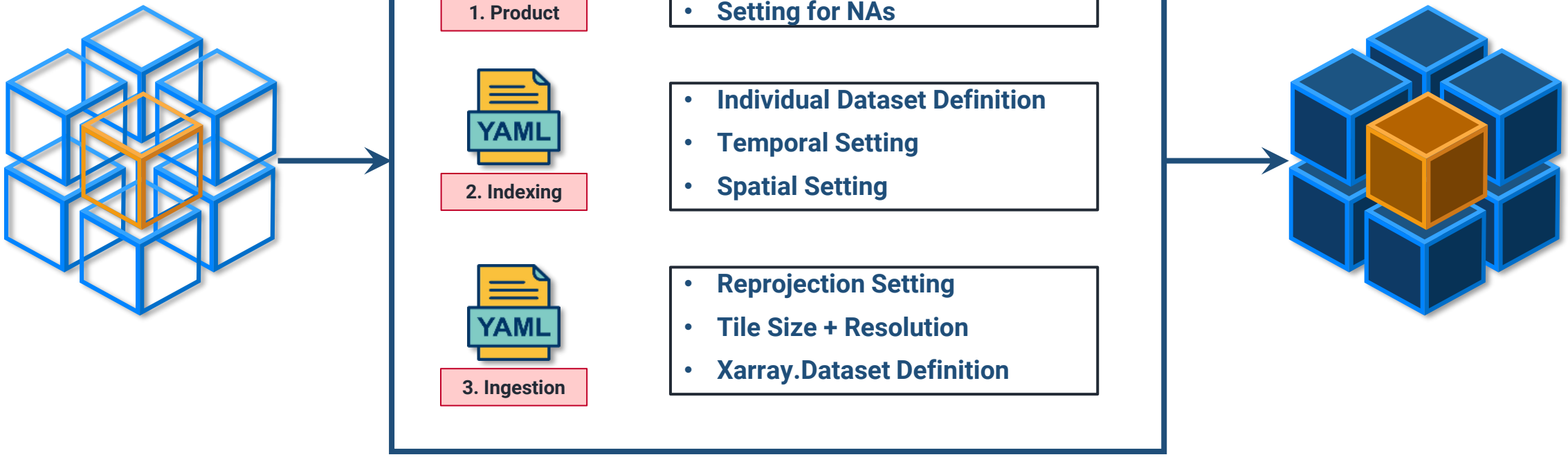
← reprojection




```
51        src_varname: 'GWP'
52        zlib: True
```

← xarray.Dataset

Ingestion YAML Summary



Main DC Commands


```
datacube <command> <argument> /path
```

1. Product

```
datacube product add /datacube/utils/product_yaml/product_eo_s2_namibia.yaml
```

```
datacube product list
> s2_l2a_namibia  Sentinel 2 ARD L2A scenes, EPSG:32733
> s2_namibia      Sentinel 2 L2A scenes processed, 10, 20 and 60 m UTM
```

2. Indexing

```
python3 /datacube/utils/index_py/prep_j2_to_yaml.py /datacube/s2_download/L2A/*.SAFE
--output /datacube/utils/index_py/indexed_yamls
```

```
datacube -v dataset add /datacube/utils/index_py/indexed_yamls/*.yaml
--product s2_namibia
```

3. Ingestion

```
datacube -v ingest -c /datacube/utils/ingest_yaml/ingest_eo_s2_namibia.yaml
 --executor multiproc 4
```