

# EUMETSAT EO Data on the Web

Making EUMETSAT EO Data Part of the Indexed Web

**Workshop on making spatial data discoverable  
through mainstream search engines**

3-4 July 2019

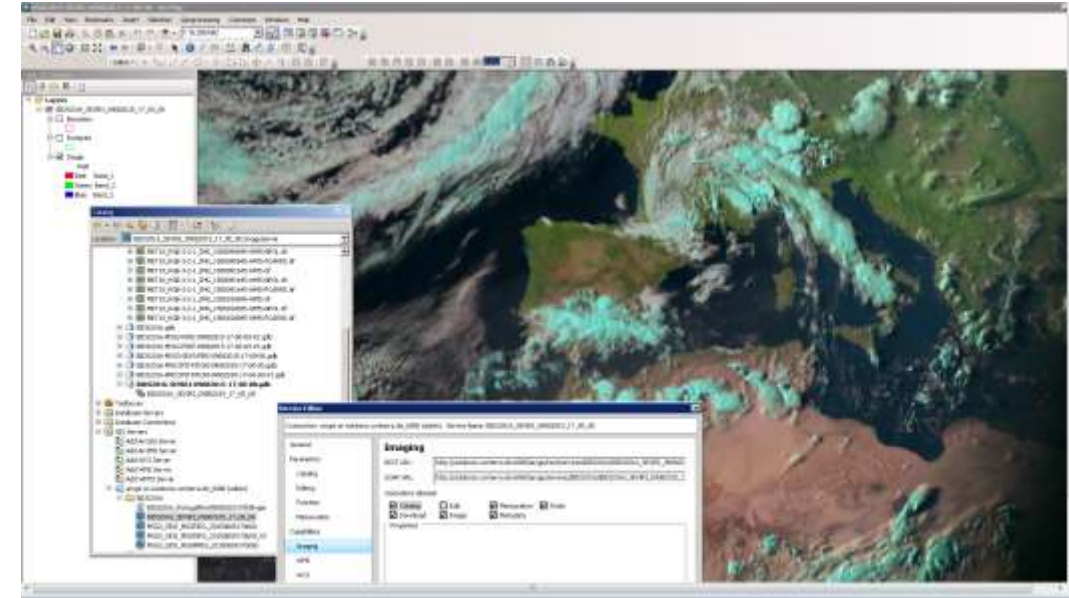
**JRC, Ispra (I)**

Dr. Uwe Voges

[u.voges@conterra.de](mailto:u.voges@conterra.de)

# Background: EUMETSAT OLDA

- Data of EO providers (like EUMETSAT) often **logically and/or physically managed as EO Collections and EO Products**
  - > **EO Collection: set of EO Products**, sharing a common specification or characteristics, e.g.:
    - > same platform (e.g. MSG) and same instrument (e.g. SEVIRI) acquired over an Area of Interest (AOI) at different times (TOI). Example EO Products: →
  - > EO Product metadata is distinguished from EO Collection metadata
- What is **OLDA (On Line Data Access)** ?
  - > Allows users and machine clients to easily and efficiently **access and download EUMETSAT products via HTTP**
- EO Collections in OLDA comprise a very **large number of products**



## Example EUMETSAT EO Products:

High Rate SEVIRI Level 1.5 –  
MSG / SEVIRI (GeoTIFF) +

Multi-Sensor Precipitation Estimate  
(mm/hr) (GRIB)

(Europe/Africa, 9.8.2015 17:00)

- > Loaded into **Esri ArcGIS** as mosaic-  
/raster-datasets and published as Image  
Services + WMS/WCS

# OLDA REST-API

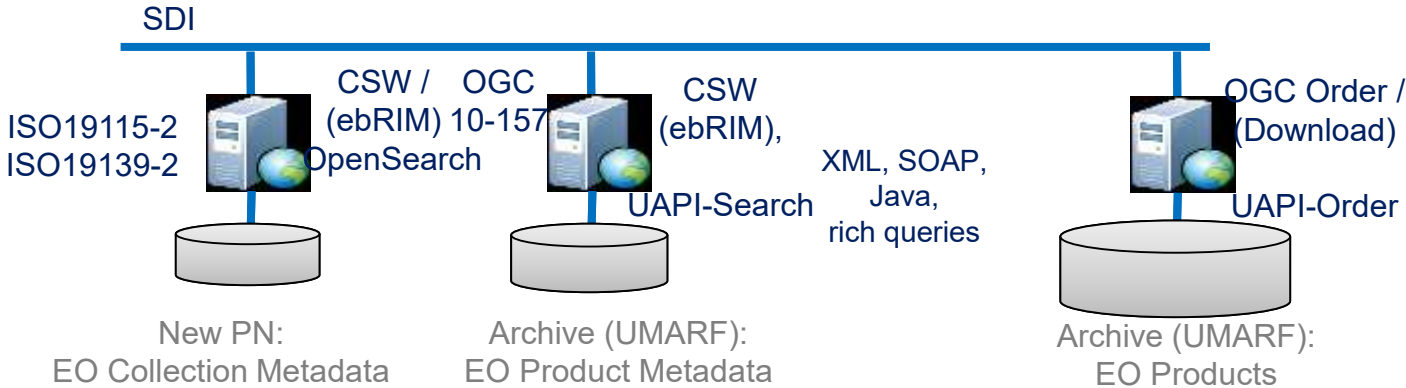
- **OLDA Data Access API** aggregates different API's:
  - > **OLDA REST API (focus here)**
  - > OLDA DOWNLOAD API (access to (parts of) the data or metadata, subscription,...) (REST)
  - > OpenSearch-EO Products (EOP Search)
  - > OpenSearch-EO Collections (Coll Search)
  - > OLDA INGESTION API (Ingestion) (REST)
- **OLDA REST API**
  - > **browsing collections and subsets of related products**
  - > to allow GUI's (usually web browsers based) and machine clients (e.g. hosted processing applications) an easy and efficient access to the data
  - > Definition in OpenAPI

# Advance OLDA REST-API to SDW (Linked Data) Services

- Further Targets: **advance REST-API to SDW (LD) Services** for EO Collections and Products
  - > Expose, search and access Collections and Products in a web-friendly way
  - > make **Collections** and **Products** discoverable through mainstream search engines
- Requirements:
  - > **interaction** using the **HTTP** protocol
  - > **persistent HTTP URIs** for **collections** and **products**
  - > **linking**: collection<->product, alternative formats, access, other metadata (e.g. satellite def.),...
  - > ensure that all resources (or subsets) can be reached via links from a “landing page”
  - > **classification of resources using vocabularies** supported by main search engines
  - > **representations** for consumption by humans (**HTML**), web-developers (**JSON**) and search engine crawlers (**HTML with annotations**)
  - > **APIs** to access data should be **self-describing**

# OLDA SDW Services - Architecture

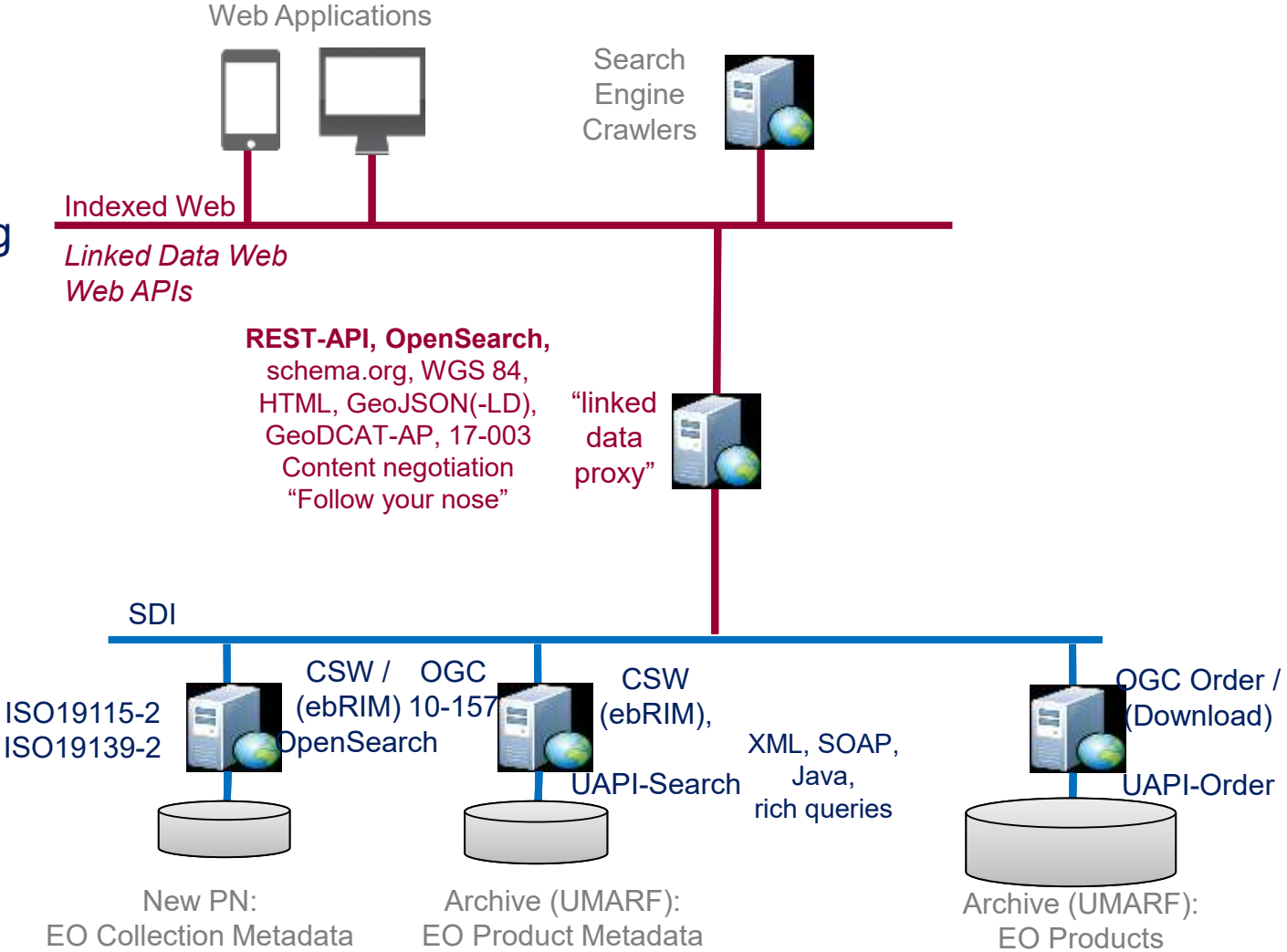
- Situation before we started...



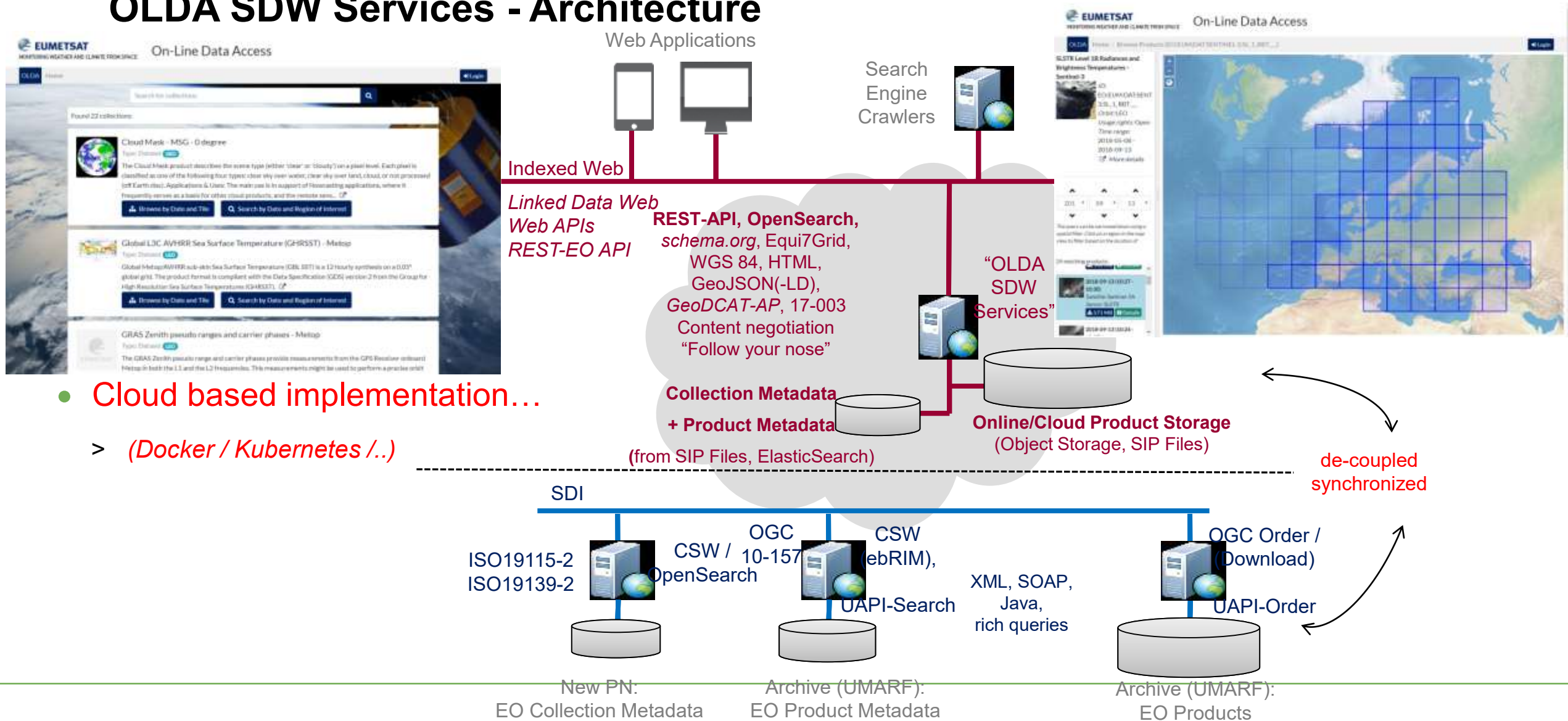
The following 3 figures are inspired by figures from: Portele et al., Spatial Data on the Web using the current SDI. Report of the research results in the Geonovum testbed "Spatial Data on the Web", 2016

# OLDA SDW Services - Architecture

- "Naive" approach...
  - > Just setting SDW/LD proxy on top of existing infrastructure
  - => not efficient



# OLDA SDW Services - Architecture



• Cloud based implementation...

> (Docker / Kubernetes /..)

# OLDA SDW Services – REST-API

## > REST-API: Browsing collections / products

- currently very basic HTML representations
- basic path: available collections (HTML + JSON) →
- selection provides metadata of the collection (derived from ISO) (HTML and JSON) →
- allows browsing through related products
- Many collections have very large number of datasets
- therefore browsing by subsets
- Subsets based on following “axis”
  - time (Year / Month / Day)
  - spatial extent (Equi7Grid: Zone + L6 grid)
- set of remaining EO products narrowed step by step by drilling into named value ranges of the axis

The screenshot displays the Pathfinder Online Data Access REST API interface. At the top, a blue header reads "PATHFINDER ONLINE DATA ACCESS REST API". Below it, a list of collections is shown, including "EO:EUM:DAT:SENTINEL-3-SR-1-SRA", "EO:EUM:DAT:MFR-CDR", "EO:EUM:DAT:MSG-CLM", "EO:EUM:DAT:MSG-MSG15-RSS", "EO:EUM:DAT:MSG-HRSEVIRI", "EO:EUM:DAT:METOP-GRARNX", "EO:EUM:DAT:MSG-HRSEVIRI-HRIT", and "EO:EUM:DAT:METOP-IASSND02". A red arrow points from the "EO:EUM:DAT:METOP-IASSND02" collection to a detailed view of that collection. This view includes the title "EO:EUM:DAT:METOP:IASSND02", the subtitle "IASI Combined Sounding Products - Metop", and a description: "The main objective of the Infrared Atmospheric Sounding Interferometer (IASI) is used for the determination of trace gases such as ozone, carbon dioxide, and methane. Parameters: IASI Ozone and IASI Trace Gases content. Total: 8 ozone amounts in thick layers, Total column N2O, CO, CH4, ...". Below the text is a "Time Range: 2008-02-13 -" and a polar projection map of the Earth showing a vertical strip of data. A color scale at the bottom of the map ranges from 200 to 320 K, labeled "Surface skin temperature (K)". To the right of the map, there are links for "Browse products" and "Show products (8)". Below this, a list of products is shown, including "W\_XX-EUMETSAT-Darmstadt-HYPERSPECT+...\_55Z-2017-12-07T10:14:54Z", "W\_XX-EUMETSAT-Darmstadt-HYPERSPECT+...\_58Z-2017-12-07T09:17:57Z", and "W\_XX-EUMETSAT-Darmstadt-HYPERSPECT+...\_58Z-2017-12-07T04:11:57Z". Further down, there are links for "Browse products" and "Show products (8)". At the bottom, there is a list of years: "07", "09", "12", and "2017".



# OLDA SDW Services – REST-API

- on lower drilling levels => link to the remaining products
  - For products, metadata (sensing time etc) is provided plus links to the (meta)data access (HTML/JSON, e.g. OGC 17-003)
- API implements HATEOAS (Hypertext as the engine of application state) with encodings:
  - HTML (for direct usage in web browser)
  - GeoJSON (for machine clients)
  - beneficial for processings with simple iterations (no search)
  - usable by Jupyter Notebook....
  - Subsets are established by Elasticsearch queries
  - via facets: e.g. years for which products exist 2017, 2018...

The screenshot displays a REST API response for the collection `EO:EUM:DAT:METOP:IASSND02`. The main content is a JSON object with the following fields:

- Collection:** `EO:EUM:DAT:METOP:IASSND02`
- Sensing start time:** `2017-12-07T08:32:55Z`
- Sensing end time:** `2017-12-07T10:14:54Z`
- Mission:** `METOPB`
- Instrument:** `IASI`
- Size:** `77868`

Below the metadata is a thumbnail image of a satellite image. At the bottom, there are links for `Download as SIP`, `EOP Metadata`, and `EOP Metadata in JSON format`. A section titled **SIP Contents** lists the following files:

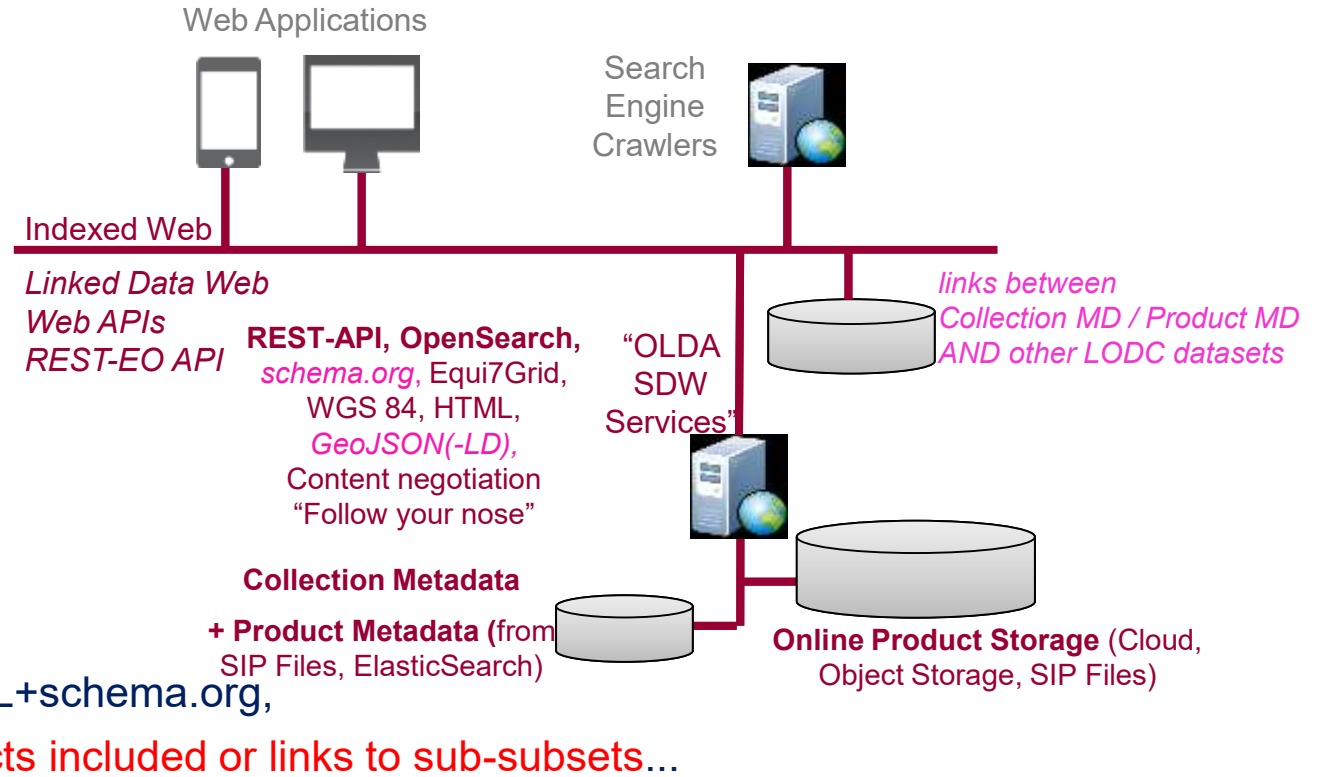
- `W_XX-EUMETSAT-Darmstadt,HYPERSPECT+S_...nc`
- `EOPMetadata.xml`
- `browse.jpg`
- `thumbnail.jpg`
- `manifest.xml`

On the right side, there is a sidebar with a breadcrumb trail: `collections / EO:EUM:DAT:METOP:IASSND02 / dates / 2017 / 12 / products`. Below the breadcrumb, there are links for `Titles`, `Show products (8)`, and a list of dates: `07`, `09`, `12`, and `2017`. Red arrows point from the text in the list to these elements in the screenshot.

# OLDA SDW Services – Pending (or working on)

- EO Collections

- HTML without schema.org annotations
- MD => proprietary model, JSON based
  - Future: OGC 17-084 based -> GeoJSON-LD for EO collection metadata
- representation of subsets to be improved
  - Missing: metadata of collection involved, HTML+schema.org, more metadata related to subset, list of products included or links to sub-subsets...



```

PATHFINDER ONLINE DATA ACCESS REST API
collections / EO.EUM.DAT.SENTINEL-3.OL_2.WFR___ / dates / 2018 / 01 / 01
Titles
Show products (208)
    
```

```

PATHFINDER ONLINE DATA ACCESS REST API
collections / EO.EUM.DAT.SENTINEL-3.OL_2.WFR___ / dates / 2018 / 01 / 01 / products
S3A_OL_2_WFR_20180101T230414_20180101T230447_20180103T101847_0033_026_172_4679_MAR_O_NT_002_SEN3(2018-01-01T23:04:13.794269Z - 2018-01-01T23:04:47.208626Z)
S3A_OL_2_WFR_20180101T230114_20180101T230414_20180103T051721_0179_026_172_4489_MAR_O_NT_002_SEN3(2018-01-01T23:01:13.794269Z - 2018-01-01T23:04:13.794269Z)
S3A_OL_2_WFR_20180101T225814_20180101T230114_20180103T051531_0179_026_172_4319_MAR_O_NT_002_SEN3(2018-01-01T22:58:13.794269Z - 2018-01-01T23:01:13.794269Z)
S3A_OL_2_WFR_20180101T225514_20180101T225814_20180103T051319_0179_026_172_4139_MAR_O_NT_002_SEN3(2018-01-01T22:55:13.794269Z - 2018-01-01T22:58:13.794269Z)
S3A_OL_2_WFR_20180101T225214_20180101T225514_20180103T051156_0179_026_172_3960_MAR_O_NT_002_SEN3(2018-01-01T22:52:13.794269Z - 2018-01-01T22:55:13.794269Z)
S3A_OL_2_WFR_20180101T224914_20180101T225214_20180103T051034_0180_026_172_3779_MAR_O_NT_002_SEN3(2018-01-01T22:49:13.794269Z - 2018-01-01T22:52:13.794269Z)
    
```

# OLDA SDW Services – Pending (or working on)

- EO products
  - HTML->schema.org annotations (the GeoJSON-LD representations we provide could be used)
    - for product metadata (typically based on O&M EOP XML (e.g. SIP)) experimental schema.org mapping
  - GeoJSON not aligned with latest OGC 17-003
  - proprietary data format: SIP
- missing or to be improved...
  - sitemaps for informing search engines about collections
  - to improve robots.txt and HTML meta tags (e.g., index, follow) to optimize the indexing of web sites
  - use of SEO tools (e.g., Google Search Console to check indexing status and optimize visibility)
  - alignment with OGC API Common / CAT 4 / STAC...