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

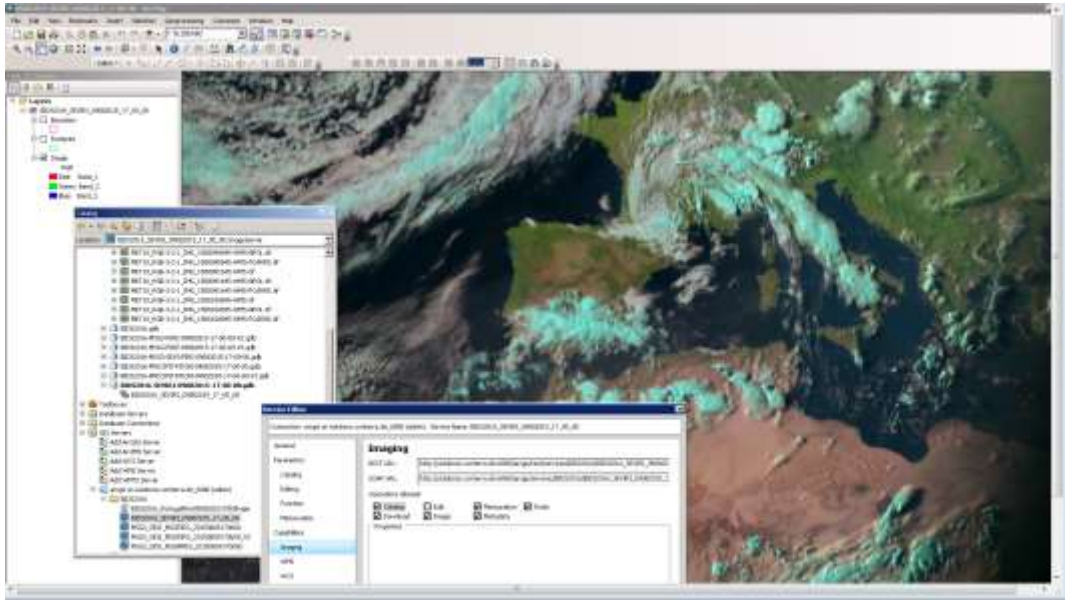
3-4 July 2019

JRC, Ispra (I)

Dr. Uwe Voges

u.voges@conterra.de

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

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

- 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

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)

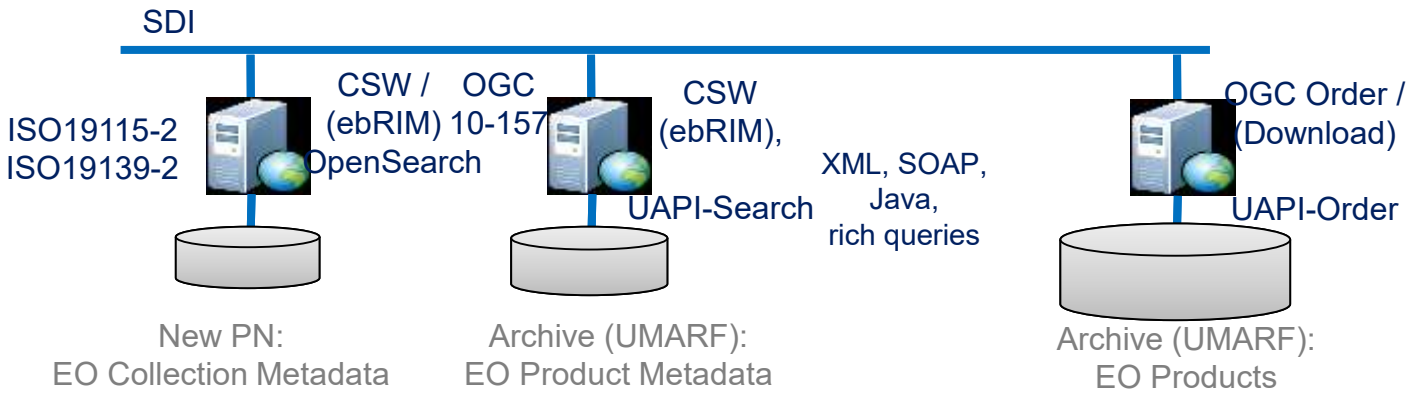
- OLDA REST API
  - > browsing collections and subsets of related products
  - > to allow GUI´s (usually web browsers based) and machine clients (e.g. hosted processing applications) an easy and efficient access to the data
  - > Definition in OpenAPI

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…

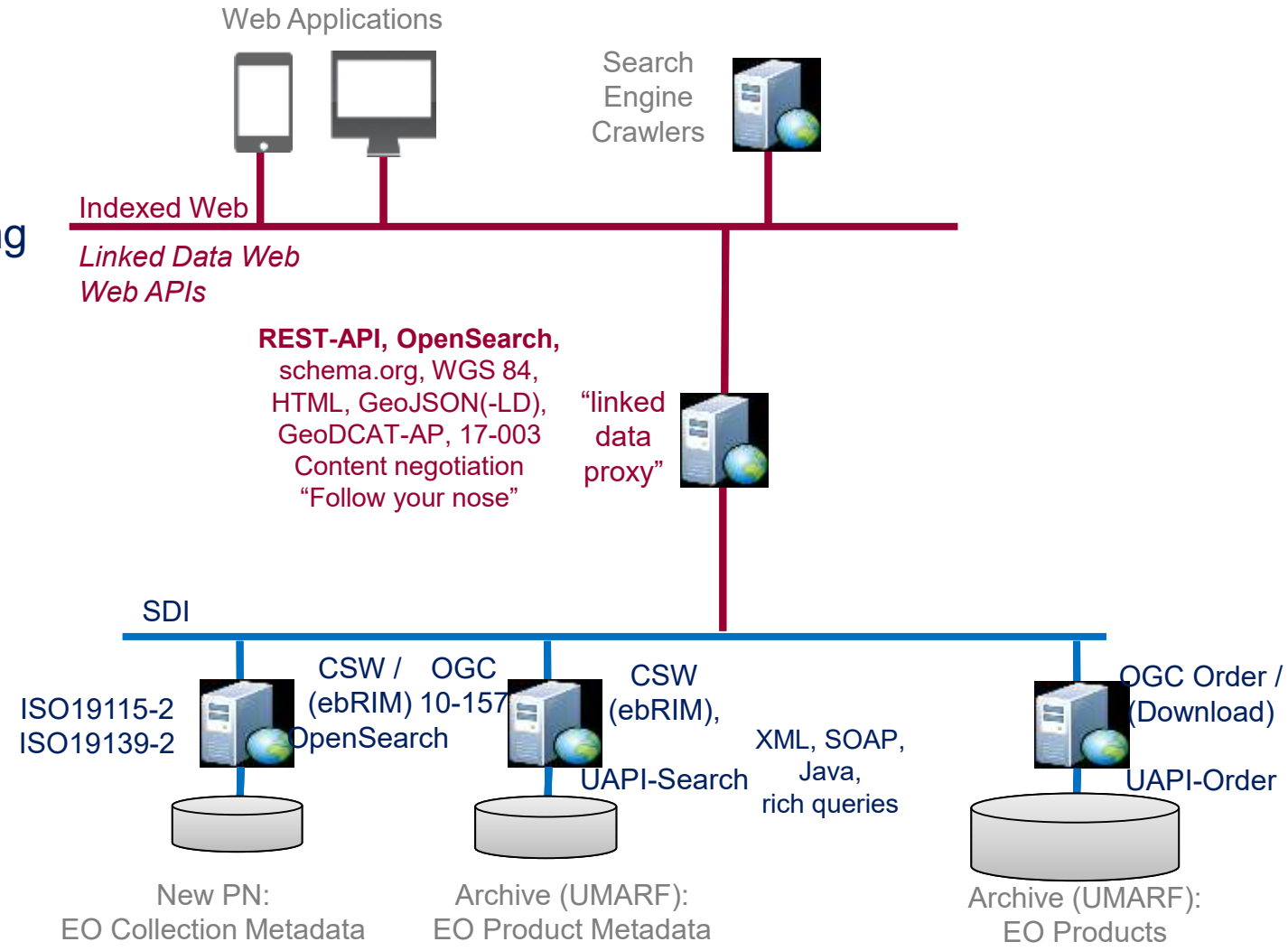


The following 3 figures are inspired by figures from: Portele et al., Spatial Data on the Web using the current SDI. Report of the research results in the Geonovum testbed “Spatial Data on the Web”, 2016

OLDA SDW Services - Architecture

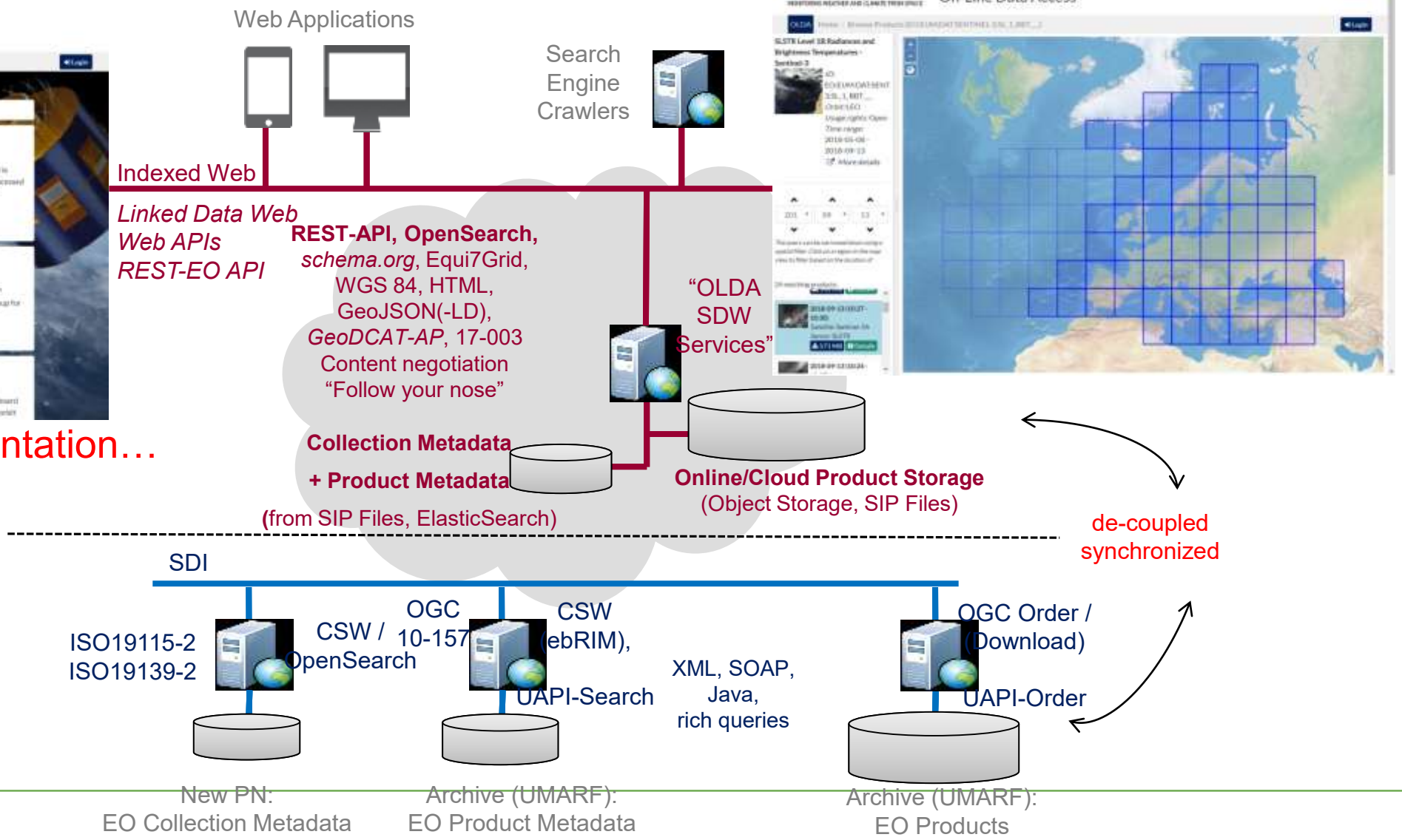
- “Naive" approach…
  - > Just setting SDW/LD proxy on top of existing infrastructure

    => not efficient



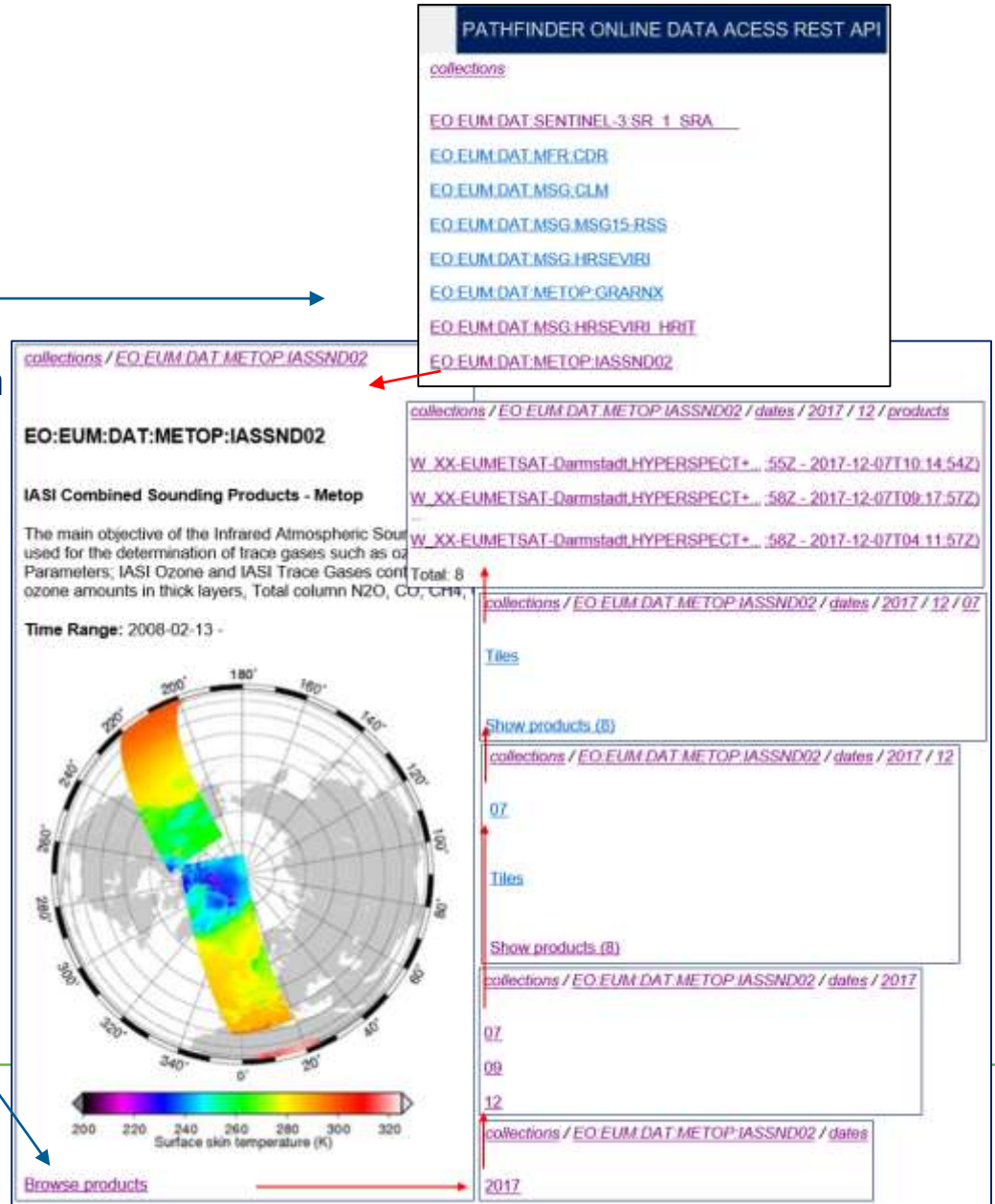
OLDA SDW Services - Architecture

- Cloud based implementation…
  - > *(Docker / Kubernetes /..)*



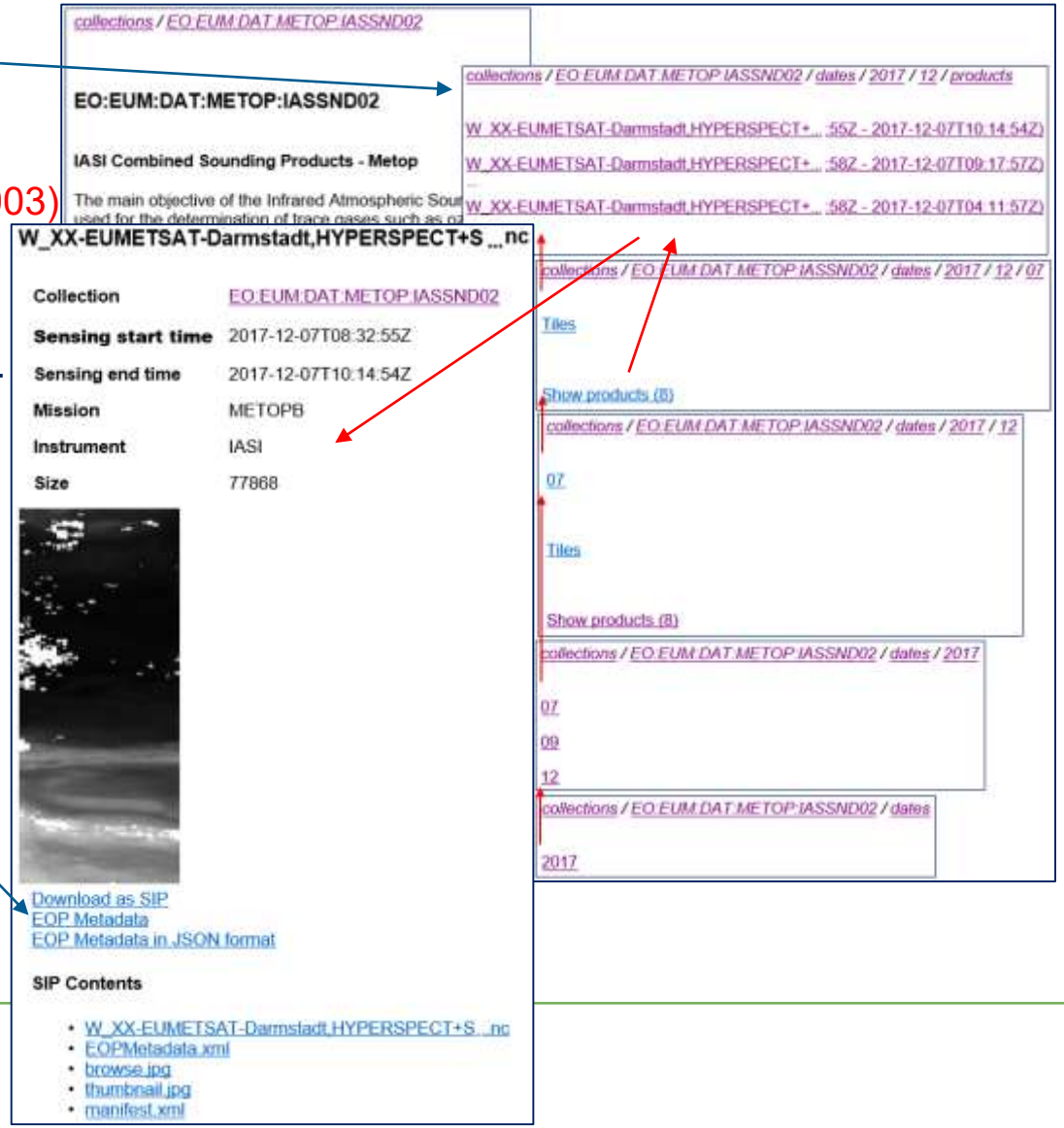
OLDA SDW Services – REST-API

- **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



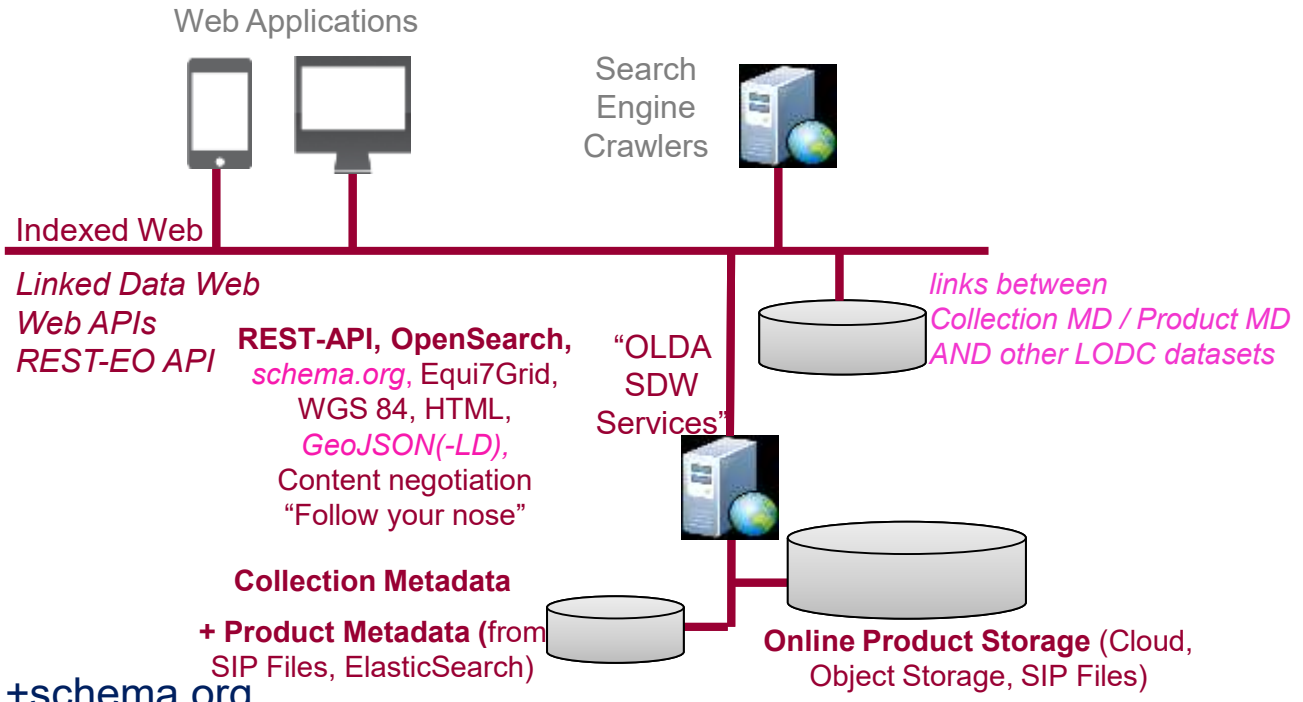
OLDA SDW Services – REST-API

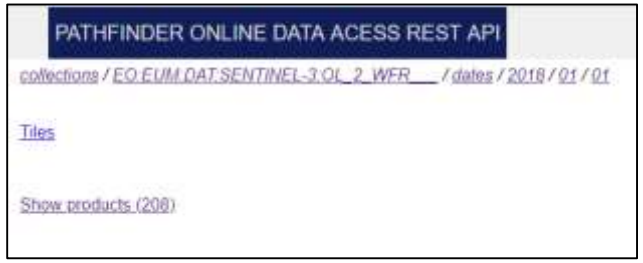
- on lower drilling levels => link to the remaining products
  - For products, metadata (sensing time etc) is provided plus links to the (meta)data access (HTML/JSON, e.g. OGC 17-003)
- API implements HATEOAS (Hypertext as the engine of application state) with encodings:
- HTML (for direct usage in web 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…



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...





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…