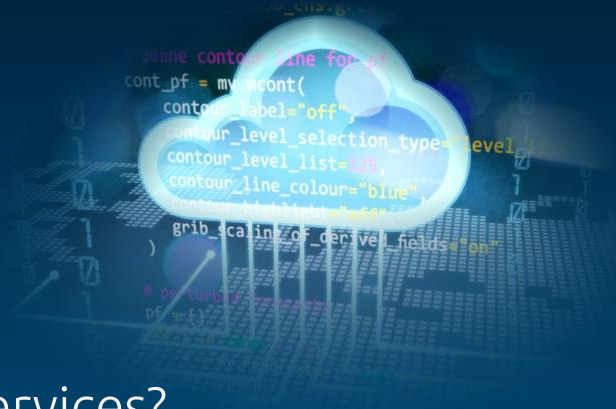
Workshop:
Weather and climate in the cloud

ECMWF

Data Tailor Web Service: cloud-native data customisation at EUMETSAT

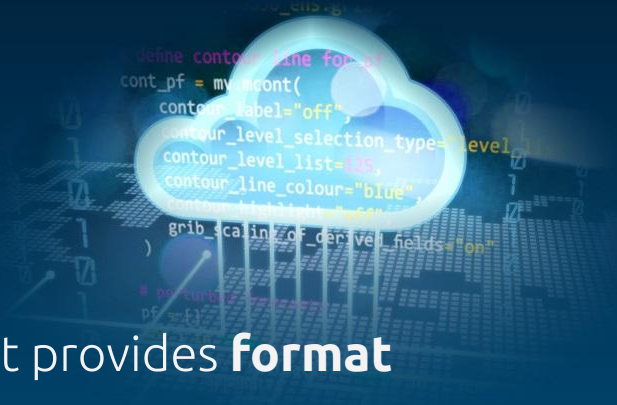
Alessandro Amici - @alexamici
B-Open - https://bopen.eu

Agenda



- What is the EUMETSAT Data Tailor?
- How does it fit in the pilot EUMETSAT Data Services?
- Integration among cloud services
  - identity federation
  - centralised authorisation and accounting
  - resources scalability
  - service-to-service communication via APIs
  - integrate complex end-user interactions across services

Data Tailor



The EUMETSAT Data Tailor is an Open Source tool that provides **format conversion** and basic **product customisation** capabilities for a set of **EUMETSAT products**

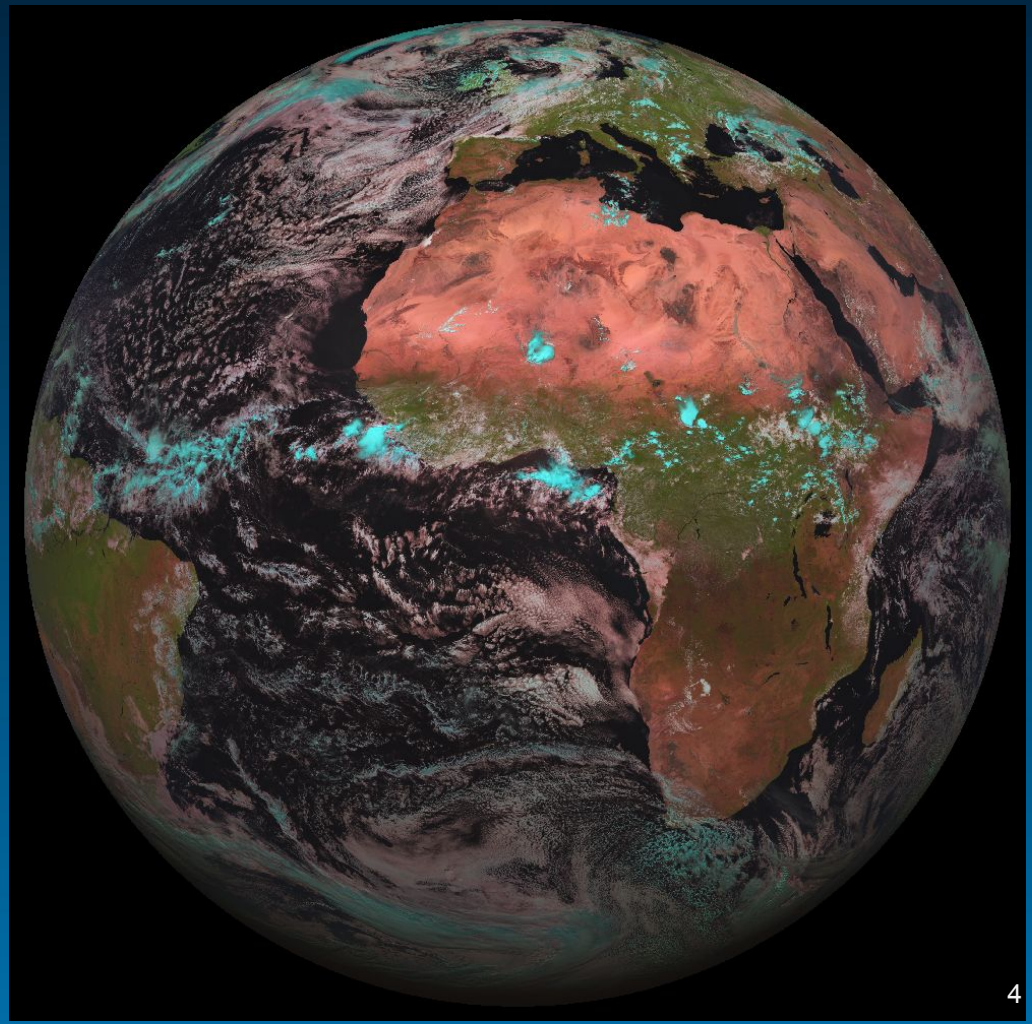
- Format conversion to well-known scientific and GIS formats
- Aggregation of native products
- Layer / band filtering
- Extraction of a region of interest (ROI)
- Reprojection and resampling

Home: https://www.eumetsat.int/data-tailor

EUMETSAT archives “native” satellite products:

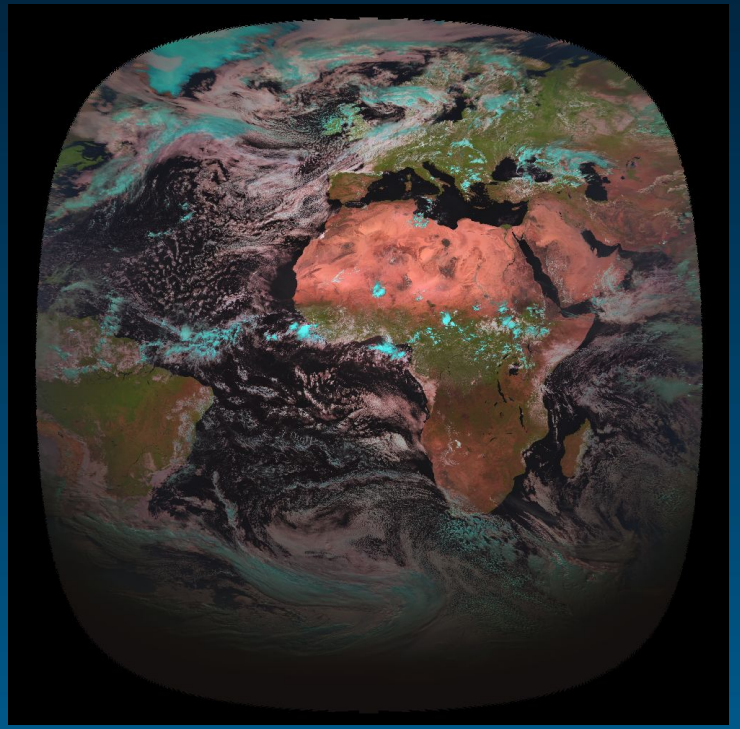
- data formats
- geometry
- granularity

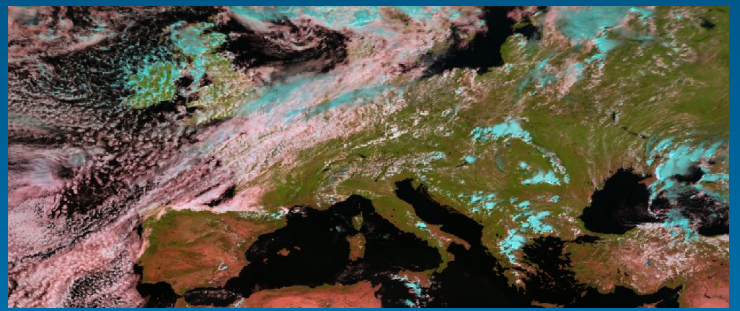
e.g. METEOSAT / SEVIRI 12-bands full-disc.



With the Data Tailor a user can:

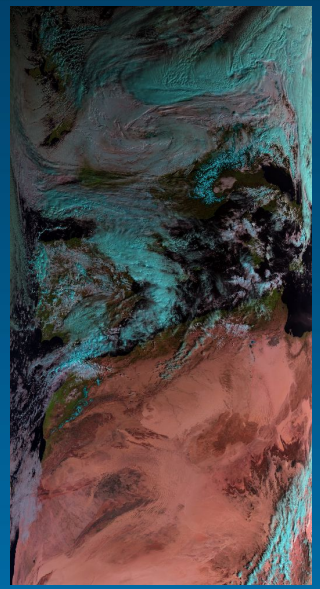
- select bands
- project to lat-lon
- extract a geographic region of interest
- convert to a scientific or GIS data format

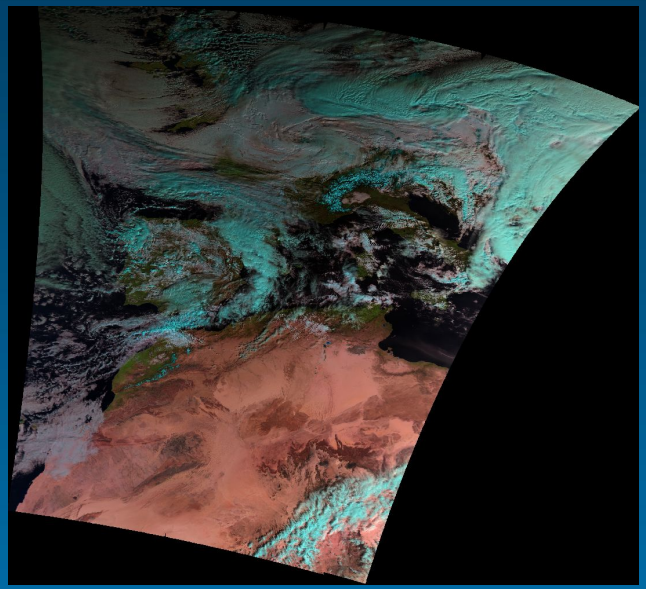


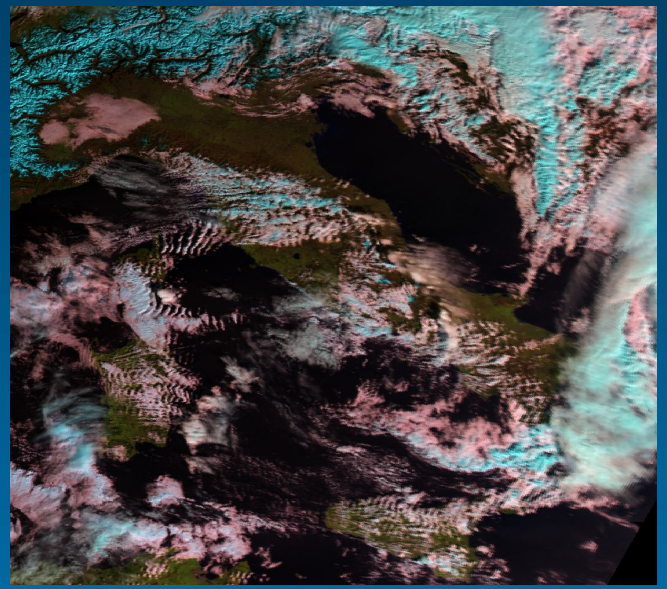


The Data Tailor can customise products of geostationary and polar satellites and higher level SAF products, e.g.

MetOp / AVHRR 6-channels







Technologies

The Data Tailor is built on top of the **Python** scientific and geospatial ecosystem:

- core: **Dask**, **Xarray**, netcdf4, NumPy, Falcon, PyJWT
- GUI: React / JavaScript
- plugins: **GDAL**, SciPy, Pandas, cfgrib / ecCodes, h5py

The Data Tailor is also one of the first EUMETSAT project that went full Open Source: https://gitlab.eumetsat.int/open-source/data-tailor

Products supported by the Data Tailor

- AMSU-A GDS Level 1B - Metop
- ASCAT Level 1 Sigma0 at Full Sensor Resolution - Metop - Global Data Service
- ASCAT GDS Level 1 Sigma0 Swath Grid - Metop
- ASCAT Soil Moisture Swath Grid in NRT - Metop
- AVHRR GDS Level 1B - Metop
- Surface Solar Radiation Data Set - Heliosat (SARAH) - Edition 2.1
- ECMWF - Near-surface air temperature
- Global L3C AVHRR Sea Surface Temperature - Metop
- GOME-2 GDS Level 1B - Metop
- Polar Multi-Sensor Aerosol Optical Properties - Metop
- GPCC - Thickness of the liquid water equivalent total precipitation
- HIRS GDS Level 1B - Metop
- High Rate SEVIRI Level 1.5 Image Data - MSG
- Scatterometer Root Zone Soil Moisture (RZSM) Data Record 10km resolution - Multimission
- IASI GDS Level 1C - All Spectral Samples - Metop
- IASI Combined Sounding Products - Metop
- LSA SAF - Climate Data Records - Fraction of absorbed photosynthetically active radiation
- LSA SAF - Climate Data Records - Fractional vegetation cover
- LSA SAF - Climate Data Records - Leaf area index
- Land Surface Temperature Climate Data Record - MSG
- LSA SAF - Climate Data Records - Reference evapotranspiration
- LSA SAF - Climate Data Records - Normalized difference vegetation index
- LSA SAF - Near-Real Time - Land surface temperature
- LSA SAF - Near-Real Time - Reference evapotranspiration
- Daily Fraction of Absorbed Photosynthetic Active Radiation - MSG
- Evapotranspiration - MSG - 0 degree
- Reference Evapotranspiration - MSG
- MHS GDS Level 1B - Metop
- Land Surface Temperature Climate Data Record - MSG
- Atmospheric Motion Vectors - MSG
- Cloud Analysis - MSG
- Cloud Mask - MSG
- Active Fire Monitoring (GRIB) - MSG
- Multi-Sensor Precipitation Estimate (GRIB) - MSG
- Optimal Cloud Analysis - MSG
- Flexible Combined Imager (FCI) Level 1C - MTG
- Flexible Combined Imager (FCI) Level 2 Global Instability Index (GII) - MTG
- IRS Level 1 Principal Component - MTG
- HRI Level 1.5 Image Data - MFG
- ASCAT Winds and Soil Moisture at 25 km Swath Grid - Metop
- ASCAT Winds at 25 km Swath Grid - Metop
- ASCAT Coastal Winds at 12.5 km Swath Grid - Metop
- ASCAT L2 25 km Winds Data Record Release 1 - Metop
- ASCAT L2 12.5 km Winds Data Record Release 1 - Metop
- ERS Scatterometer L2 25 km Winds Data Record Release 1 - ERS
- SeaWinds L2 25 km Winds Data Record Release 1 - QuikSCAT

DTWS in EUMETSAT Data Services

Pilot EUMETSAT Data Services:

- EUMETSAT Data Store - https://data.eumetsat.int
  - catalogue and archive
- EUMETSAT Data Tailor - https://tailor.eumetsat.int
  - product customisation service
- EUMETView - https://view.eumetsat.int
  - online map service

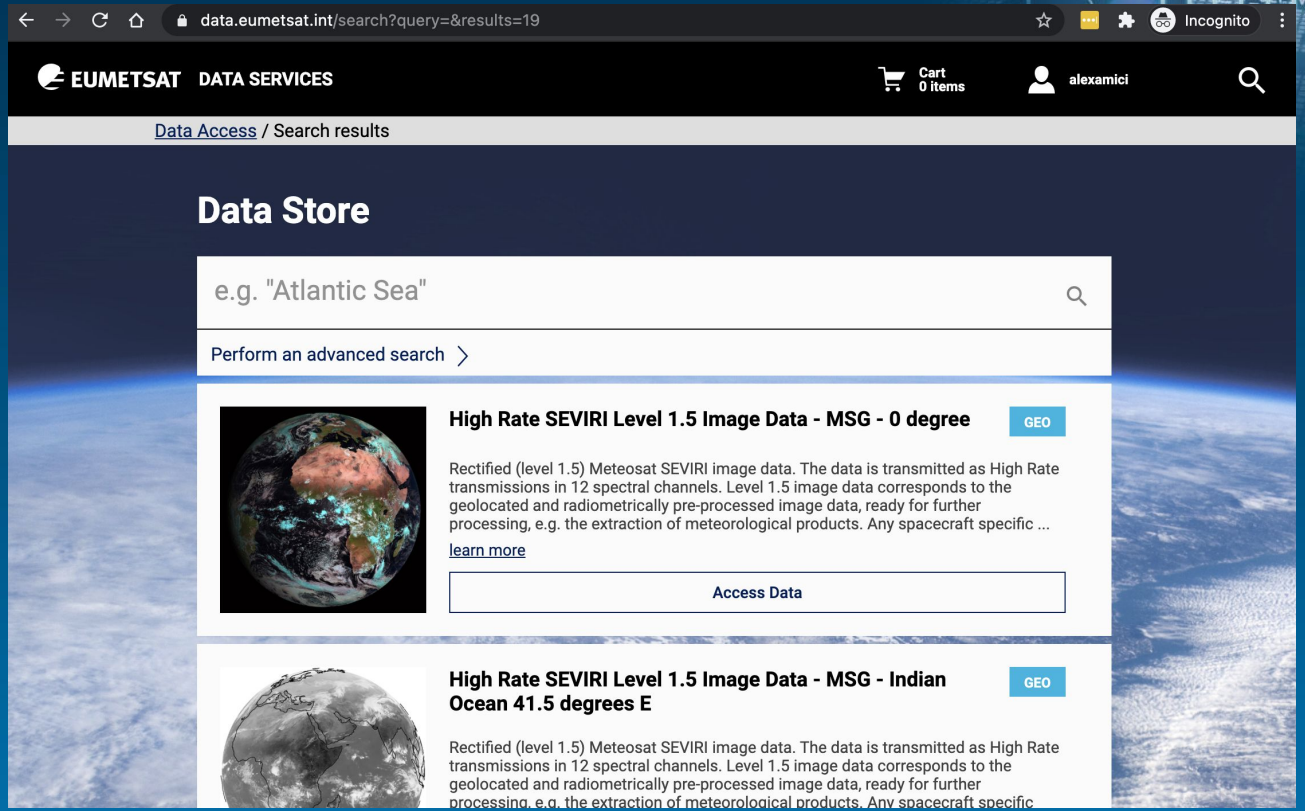
EUMETSAT

Data Tailor Web Service

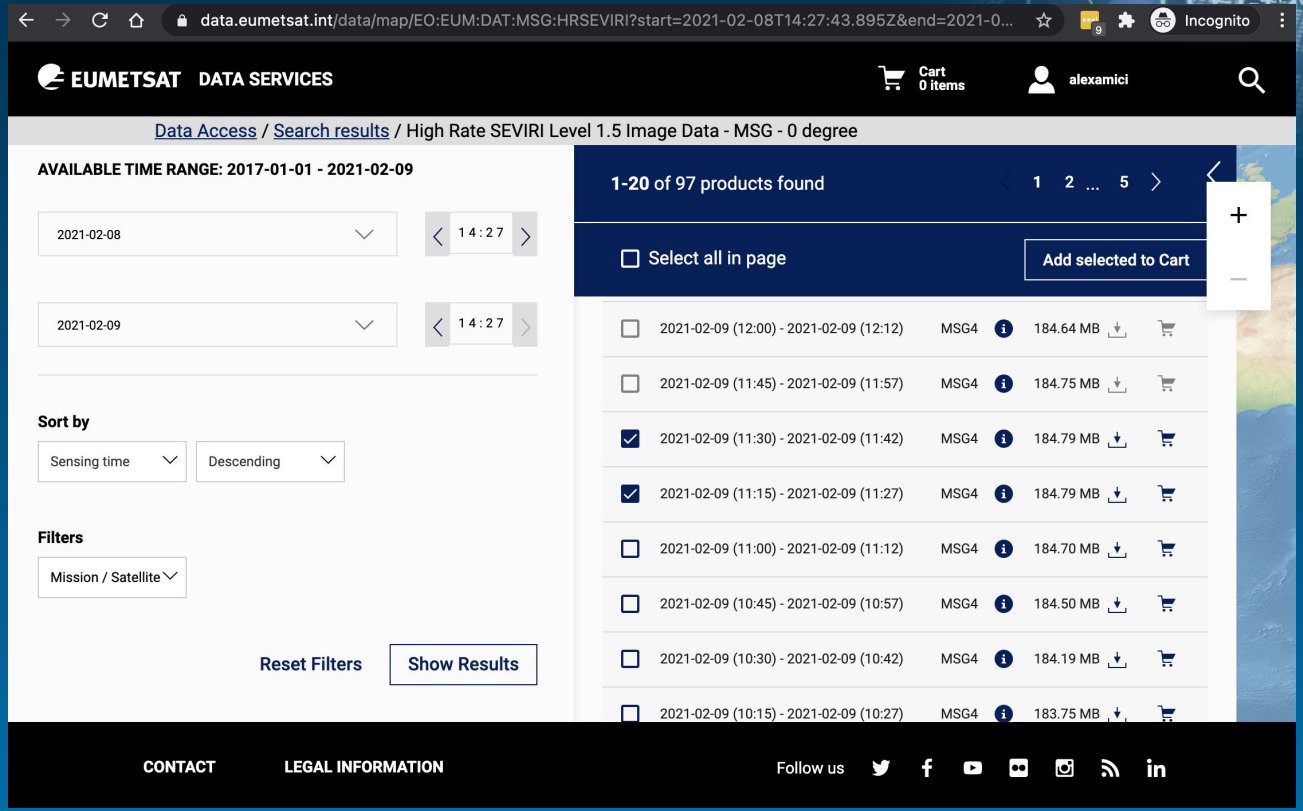
The Data Tailor Web Service forms part of the pilot EUMETSAT Data Services and is the adapter between the Data Store and the end-users:

- download only the data they need (bands and spatial area)
- download ready-to-use data (data type, projection and format)

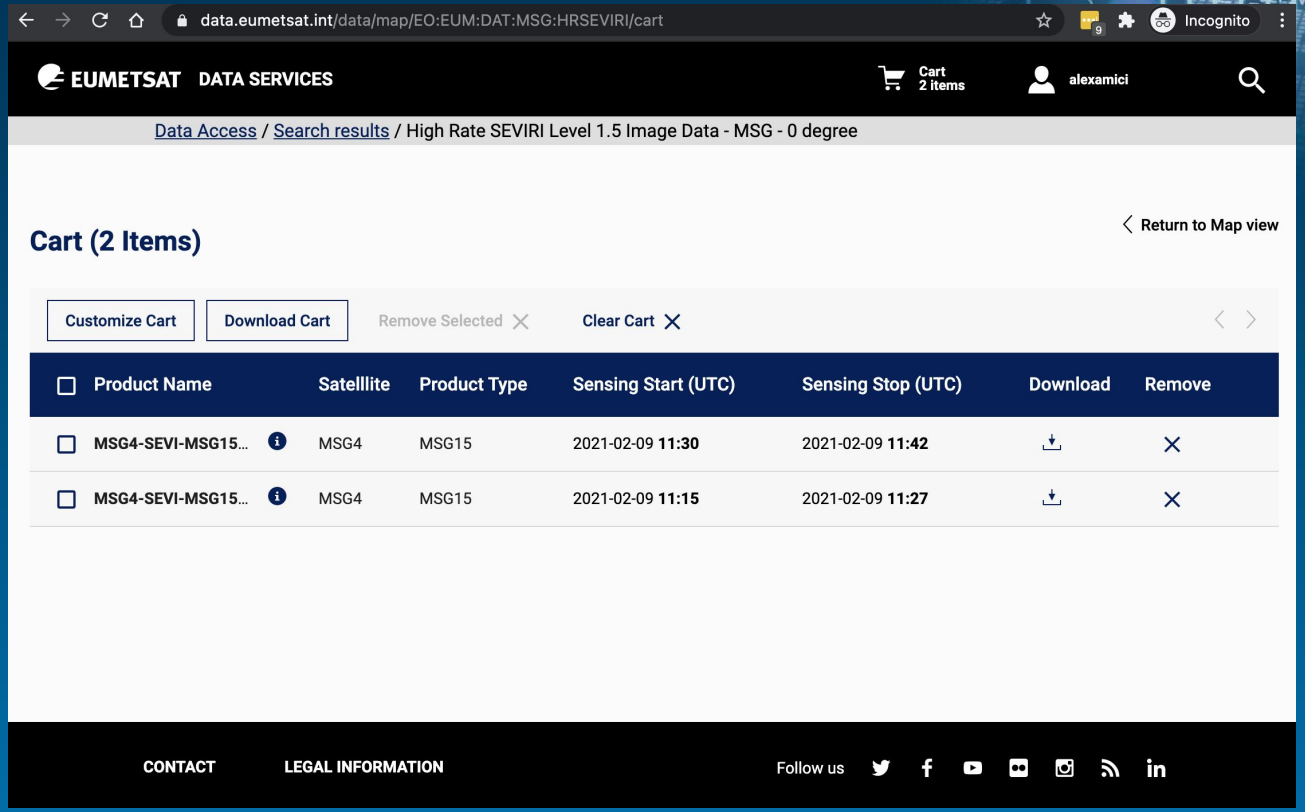
Pilot Data Services: user journey 1/6



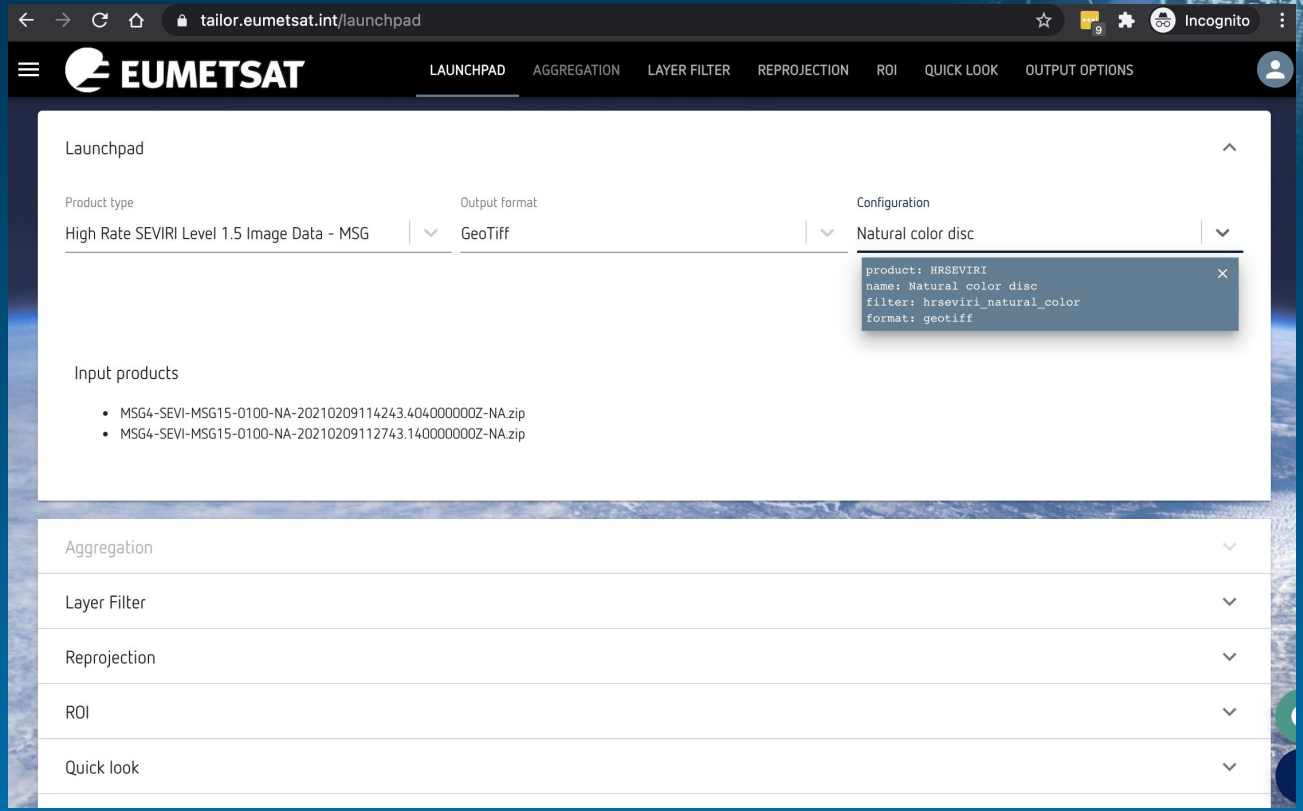
Pilot Data Services: user journey 2/6



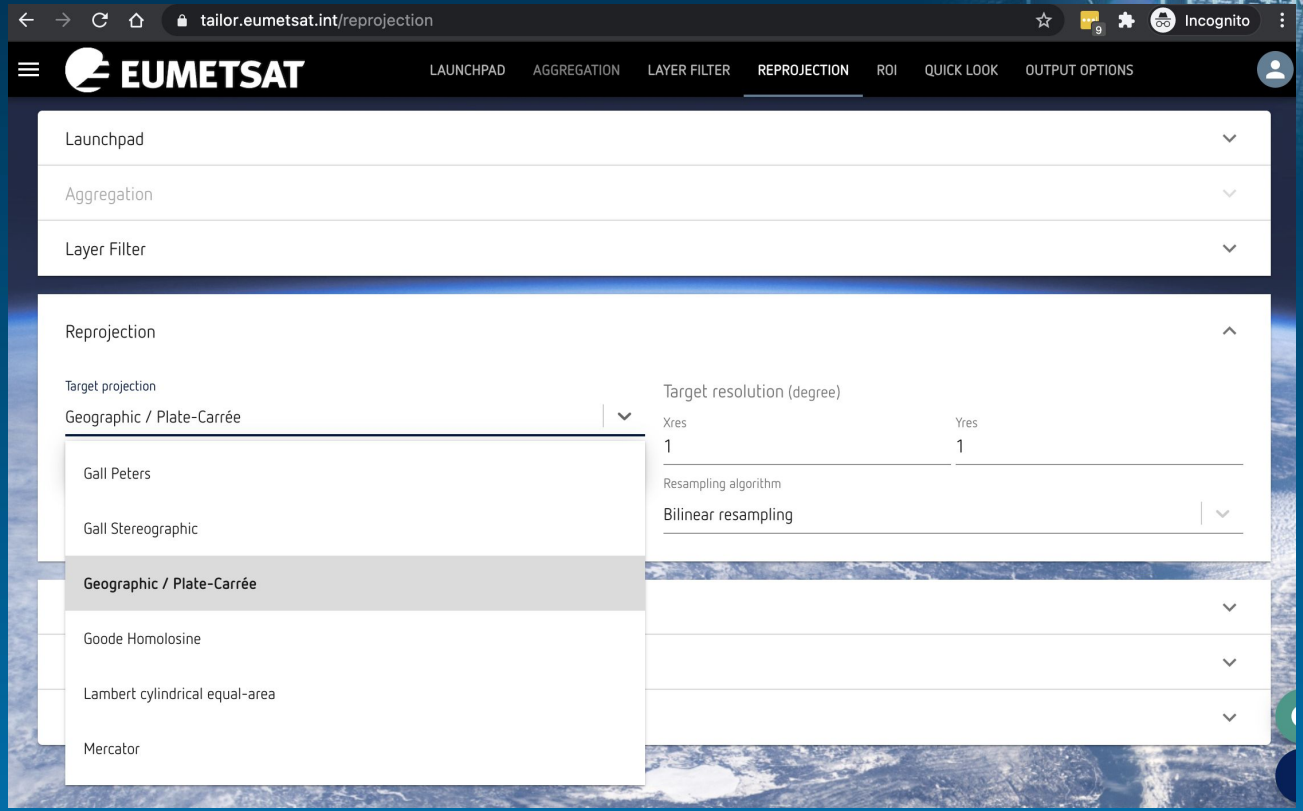
Pilot Data Services: user journey 3/6



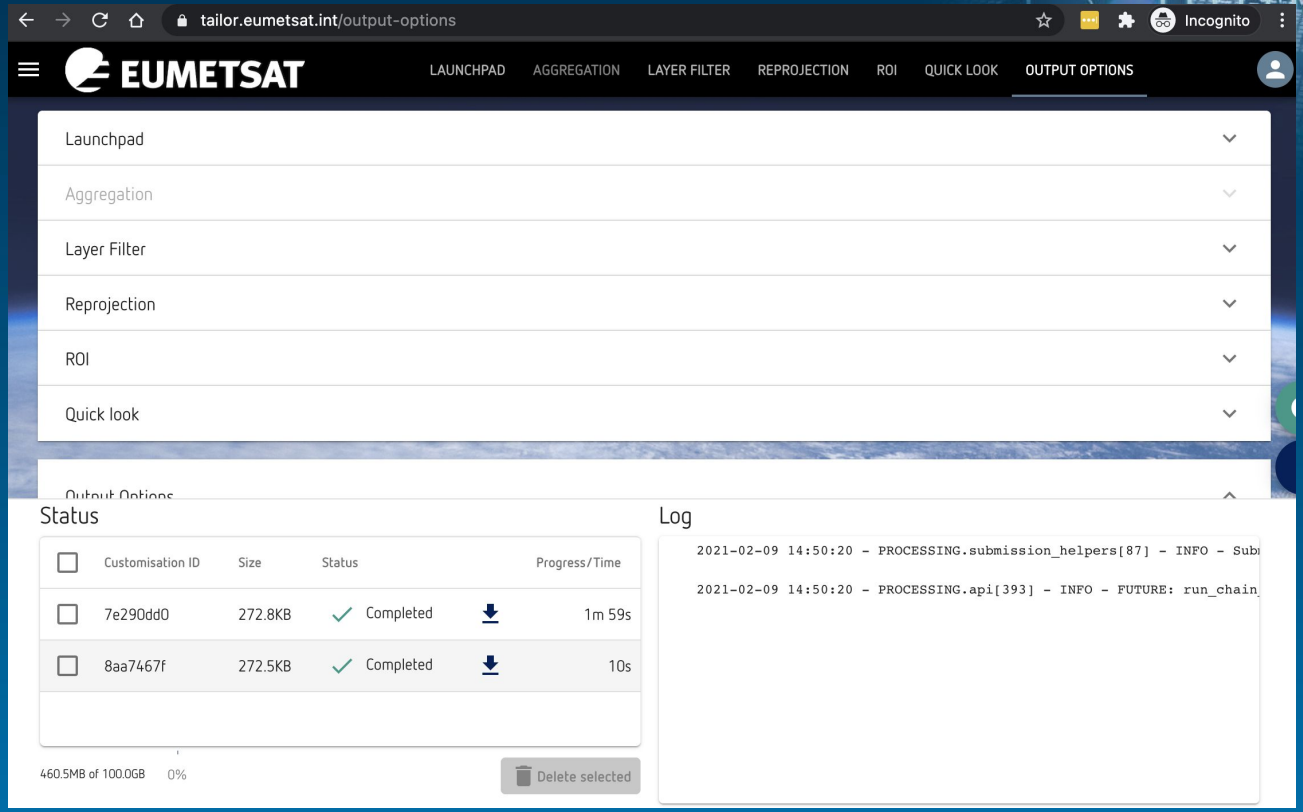
Pilot Data Services: user journey 4/6



Pilot Data Services: user journey 5/6

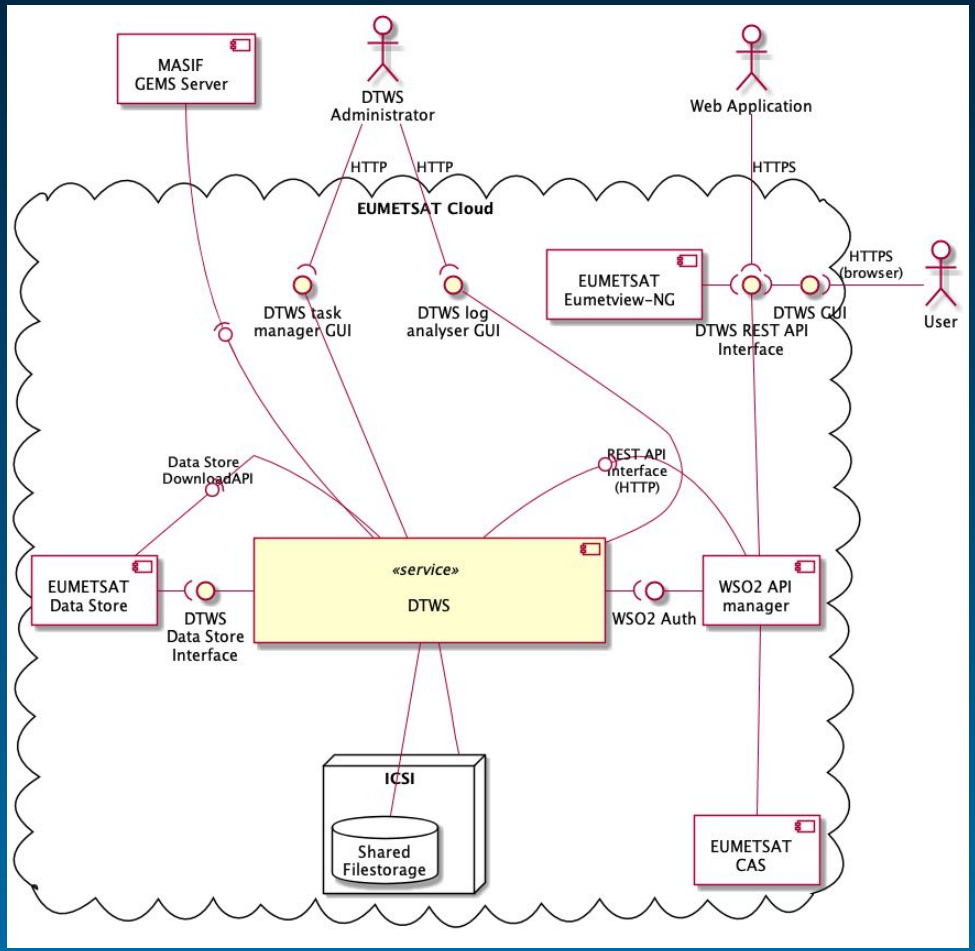


Pilot Data Services: user journey 6/6



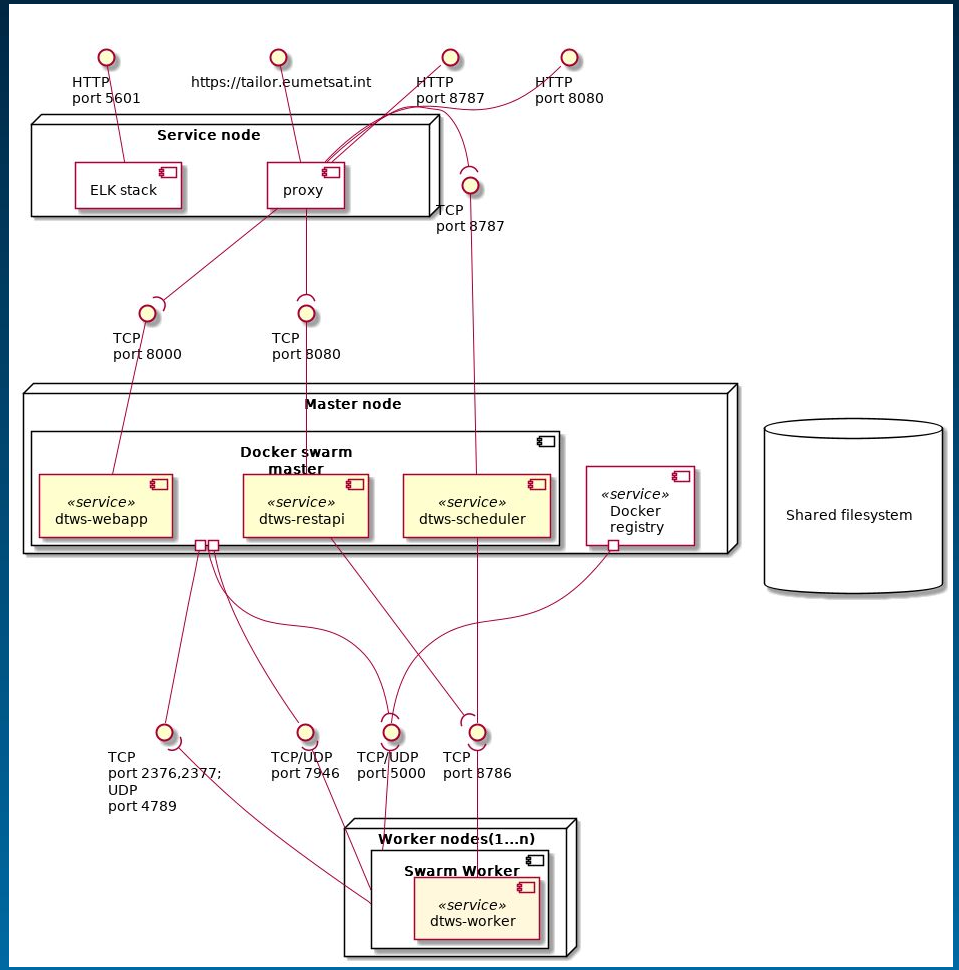
Internal cloud services:

- scalable computing and storage (ICSI)
- monitoring service (MASIF)
- identity service (CAS)
- API management service (WSO2)
  - authorisation and accounting
  - OpenAPI



Data Tailor Web Service architecture:

- dtws-webapp (JS, React)
- dtws-restapi (Python, Falcon)
- dtws-scheduler (Dask)
- dtws-worker (Docker, Dask, Data Tailor)



Cloud integration: AAA

OAuth unify Authentication, authorisation and accounting:

- user logs into the Data Store GUI and search the catalogue
  - authentication via CAS, API authorisation and accounting from WSO2
- Data Store directs the user to the Data Tailor GUI
  - identity passes transparently to the DTWS (CAS)
  - DTWS accesses the native products as the user (WSO2)
  - user configuration, processing and storage quotas and access to customised products are authorised and accounted by the DTWS* (room for improvement)

Cloud integration: computing

- Scalable pool of VMs (ICSI)
- Orchestration via dedicated Docker Swarm* (will migrate to ICSI Kubernetes Service):
  - dedicated Dask scheduler
  - scalable pool of Dask workers
  - DTWS maintains per-user processing quotas
- Customisation are neither cached nor shared between users*

Cloud integration: storage

- Scalable NFS share (ICSI)
- DTWS manages users workspaces*:
  - DTWS maintains per-user product storage quotas

Cloud integration: GUI

The Data Store and the Data Tailor Web Service have dedicated GUIs

- Single sign on experience
- Passing user configuration between services
  - Data Store selection
- Crossing GUI boundaries*
  - look and feel, widget set
  - cart / download experience not integrated with the Data Store

Thank you

B-Open - https://bopen.eu

Alessandro Amici - @alexamici

Key persons: Daniel Lee (EUMETSAT) and Maurizio Bottaccio (B-Open)

EUMETSAT Data Tailor: https://www.eumetsat.int/data-tailor

EUMETSAT Data Services to go operational in 2021!