# QA4EO Interoperability Infrastructure for Cal Val Activities

Erminia De Grandis[1], Fabrizio Niro[1], Massimo Cardaci[1], Giancarlo Rivolta[2], Giovanni Sabatino[2], Gabriele Brizzi[1], Constant Mazeran[3] , Daniele Di Erasmo[1], Laura Fioretti[4], Georgia Doxani[1], Alessandro Di Bella[1], David Fuertes[5], Juan Carlos Antuña-Sánchez[5]

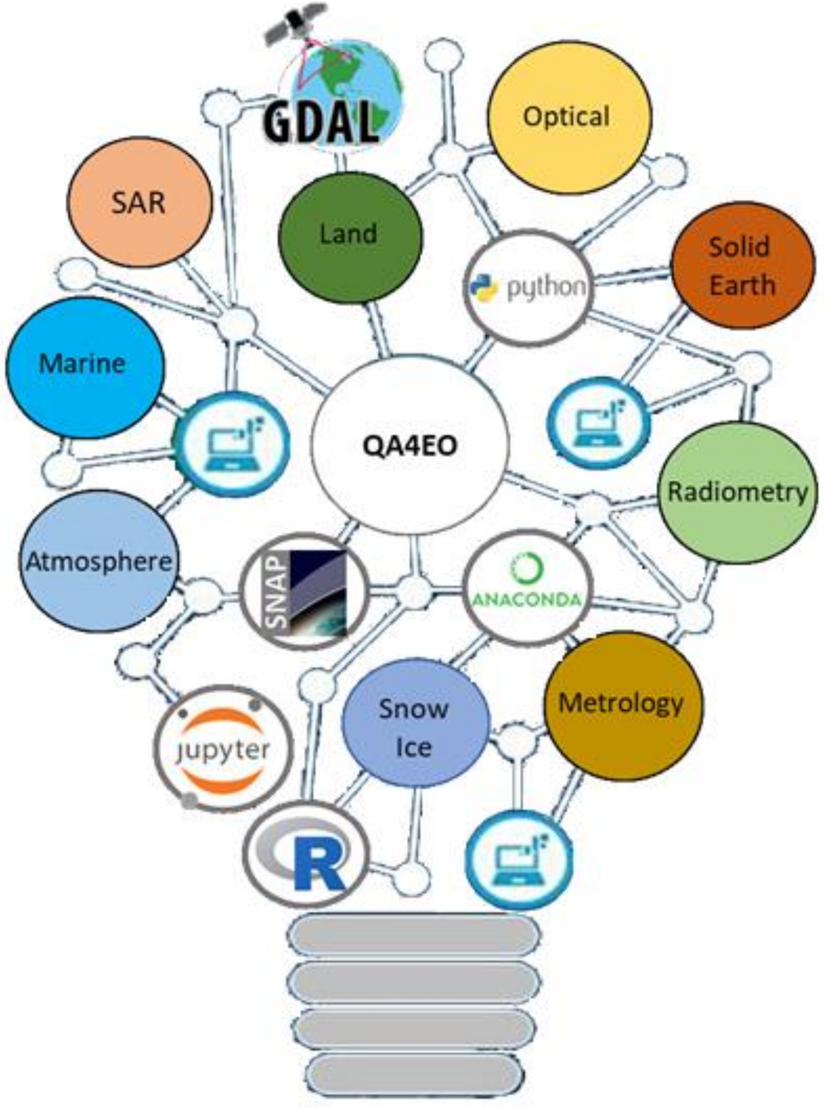
*1 Serco, Frascati (Italy) – 2 Progressive Systems, Frascati (Italy) – 3 Solvo, Antibes (France) – 4 Aresys, Milan (Italy) – 5 GRASP SAS, Villeneuve d'Ascq (France)*

**IDEAS-QA4EO Cal/Val Workshop#3 - March 31, 2022 - ESA/Esrin**

serco Progressive EXPRESS YOUR GENIUS esa solvo aresys Advancing remote sensing GRASP

# Concept






EO data access and use

## Needs

- ✓ Enhance the access and use of EO data within IDEAS-QA4EO Service activities
- ✓ Setting-up of the environment for new users
- ✓ Improve the **processing** phase
- ✓ **Collaborative environment** where to share code, results and documentation

## QA4EO Cal/Val Infrastructure

New working environment to support data provisioning, accessing and processing and where the end-users can benefit of a **scalable and cloud computing** infrastructure. Direct access to the **EO data** archive including ground-truth reference data for validation analysis

# Project Overview

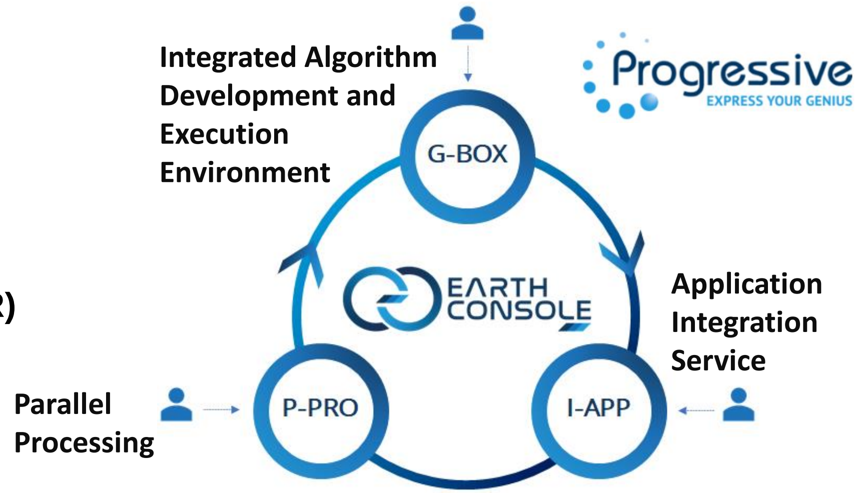
IDEAS-QA4EO

## Infrastructure Solution

**EarthConsole® (Progressive Systems):** cloud-based platform hosted on DIAS infrastructure that offers three main services used together to support all the phases of development, validation, data processing and analysis

https://earthconsole.eu/discover/

All Earth Console services are available via the ESA **Network of Resources (NoR)**




Pilot Environment
**Earth Console** – GBOX IDE Type A
4 CPU - 16 GB RAM – 64 GB SSD
1TB STORAGE
Multi Users
CREODIAS archive

| | |
|---|---|
| Aug 2020 | **Survey** and feedbacks analysis User and System **Requirements** definition |
| Nov 2020 | **Start** Infrastructure solution implementation Pilot use cases definition |
| Feb 2021 | Implementation of a working **Pilot** |
| Nov 2021 | Pilot **Extension** Phase |
| Mar 2022 | **Proposal** submission operational project |
| Apr 2022 | **Start** of **operational** project |

IDEAS-QA4EO

# Objectives

→ **Identification of** all possible **functionalities** of the selected cloud environment by involving different QA4EO domains to enrich the experience on the Earth Console platform

→ Use cases to demonstrate:

- Breaking the barrier of data availability
- Effective data access
- Advantages of a ready-to-use working environment
- Interoperability
- Processors and Algorithms integration
- Upload of external datasets

***Overview of the achieved results and potential of the infrastructure***

# Outcomes (I) – Break the barrier of data availability

IDEAS-QA4EO

## Search on CREODIAS collocations for S5P

Tool to apply a screening on the Sentinel 5P TROPOMI products using the CREODIAS interactive EO Data Finder.

Search collocations of S5P for a specific geolocation.

CREODIAS search criteria based on an input polygon defined in a shape file. Python script adapted by Daniele Di Erasmo.

In [8]:

```
## Import libraries
%reload_ext autoreload
%autoreload 2
import requests
import json
import yaml
import io
import rasterio
import matplotlib.pyplot as plt
from lib_dir.query_catalogue import query_catalogue
from rasterio.windows import Window
from pyproj import Proj
from math import floor, ceil
```

## Input Parameters

In [8]:

```
collection='Sentinel5P'

processingLevel='LEVEL2'  ## Processing Level for Sentinel 5P ['LEVEL1B', 'LEVEL2']

timeliness='Offline'

productType='L2__NO2___'

####
#import line vertex from csv file
polyDf = pd.read_csv('/application/earthuser5/shared/cropPoly_Roma.csv',header=0)
polyDf.head()

##################################
x1=str(polyDf['X'].to_list()[0])
y1=str(polyDf['Y'].to_list()[0])
x2=str(polyDf['X'].to_list()[1])
y2=str(polyDf['Y'].to_list()[1])
x3=str(polyDf['X'].to_list()[2])
y3=str(polyDf['Y'].to_list()[2])
x4=str(polyDf['X'].to_list()[3])
y4=str(polyDf['Y'].to_list()[3])

polygon_box='POLYGON(('+x1+' '+y1+','+x2+' '+y2+','+x3+' '+y3+','+x4+' '+y4+','+x1+' '+y1+'))'

# Search for Sentinel-5p products

result = search(
        polygon=polygon_box,
        begin_ts='2021-01-02T00:00:00.000Z',
        end_ts='2021-02-02T23:59:59.999Z',
        product=productType,
        processing_level='L2',
        processing_mode='Offline')
```

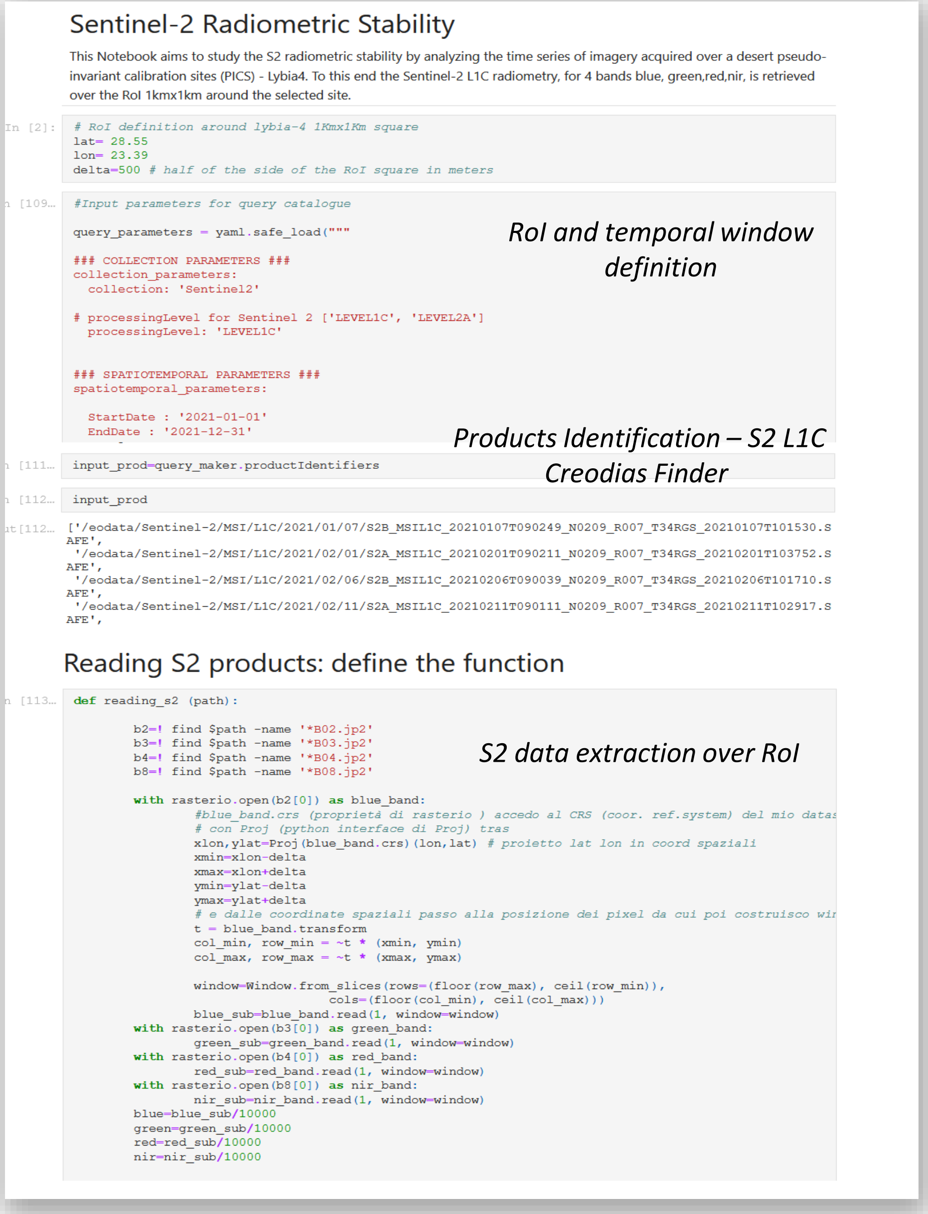

| Datasets | Products | Instrument | Locally Held |
|---|---|---|---|
| Sentinel-1A & Sentinel-1B | GRD | SAR C-BAND | Full archive |
| | RTC | | |
| | OCN | | |
| | RAW | | Last 6 months |
| | SLC | | Europe: full archive<br>Last 6 months / orderable |
| Sentinel-2A & Sentinel-2B | L1C | MSI | Full archive |
| | L2A | | Orderable<br>Cached |
| Sentinel-3A & Sentinel-3B | L1 SLSTR | SLSTR | Full archive |
| | L1 OLCI | OLCI | |
| | L1 SRAL | SRAL | |
| | L2 SLSTR (LST/WST) | SLSTR | |
| | L2 OLCI | OLCI | |
| | L2 SRAL | SRAL | |
| Sentinel-5P | L1B | TROPOMI | Full archive |
| | L2 | | |
| Sentinel-6A | L1, L2 | POS-4, AMR-C | Full archive |
| Landsat-5 | L1G, L1T, L1GT | TM | Coverage of Europe (1984-2011) |
| Landsat-7 | L1G, L1T, L1GT | ET | Coverage of Europe (1999-2017) |
| Landsat-8 | L1T, L1GT | OLI, OLI TIRS | Coverage of Europe |
| Envisat | L1 | MERIS | Global (2002-2012) |
| SMOS | L1B, L1C, L2 | MIRAS | Global (2010-present) |
| S2GL | - | - | Coverage of Europe (2017) |

| Datasets | Products | Data access |
|---|---|---|
| CAMS (Atmosphere) | All collections | Accessible over S3 or NFS protocol |
| CEMS (Emergency) | All collections | Accessible over S3 or NFS protocol |
| CLMS (Land) | All collections | Accessible over S3 or NFS protocol |
| CMEMS (Marine) | All collections | Accessible over S3 or NFS protocol |

| Datasets | Products | Instrument | Estimated locally held | Data access |
|---|---|---|---|---|
| DEM | Mapzen | SRTM, others | Aggregation of several DEM sources | Accessible over S3 or NFS protocol |
| | SRTM | SRTM | Global (56S, 60N, 180W, 180E)<br>February 2000 | Accessible over S3 or NFS protocol |
| | Copernicus | Multiple | Aggregation of several DEM sources | |
| Jason-3 | Altimeter | Altimeter | Since mission began | Accessible over S3 or NFS protocol |

https://creodias.eu/data-offer

# Outcomes (II) – Effective data access



## Sentinel-2 Radiometric Stability

This Notebook aims to study the S2 radiometric stability by analyzing the time series of imagery acquired over a desert pseudo-invariant calibration sites (PICS) - Lybia4. To this end the Sentinel-2 L1C radiometry, for 4 bands blue, green,red,nir, is retrieved over the RoI 1kmx1km around the selected site.

```
In [2]:
# RoI definition around lybia-4 1Kmx1Km square
lat= 28.55
lon= 23.39
delta=500 # half of the side of the RoI square in meters
```

*RoI and temporal window definition*

```
#Input parameters for query catalogue

query_parameters = yaml.safe_load("""

### COLLECTION PARAMETERS ###
collection_parameters:
  collection: 'Sentinel2'

# processingLevel for Sentinel 2 ['LEVEL1C', 'LEVEL2A']
  processingLevel: 'LEVEL1C'


### SPATIOTEMPORAL PARAMETERS ###
spatiotemporal_parameters:

  StartDate : '2021-01-01'
  EndDate : '2021-12-31'
```

*Products Identification – S2 L1C Creodias Finder*

```
input_prod=query_maker.productIdentifiers
```

```
input_prod
```

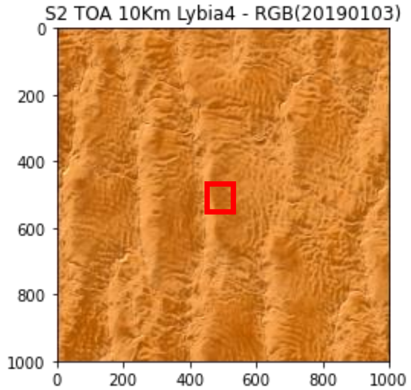
```
['/eodata/Sentinel-2/MSI/L1C/2021/01/07/S2B_MSIL1C_20210107T090249_N0209_R007_T34RGS_20210107T101530.SAFE',
 '/eodata/Sentinel-2/MSI/L1C/2021/02/01/S2A_MSIL1C_20210201T090211_N0209_R007_T34RGS_20210201T103752.SAFE',
 '/eodata/Sentinel-2/MSI/L1C/2021/02/06/S2B_MSIL1C_20210206T090039_N0209_R007_T34RGS_20210206T101710.SAFE',
 '/eodata/Sentinel-2/MSI/L1C/2021/02/11/S2A_MSIL1C_20210211T090111_N0209_R007_T34RGS_20210211T102917.SAFE',
```

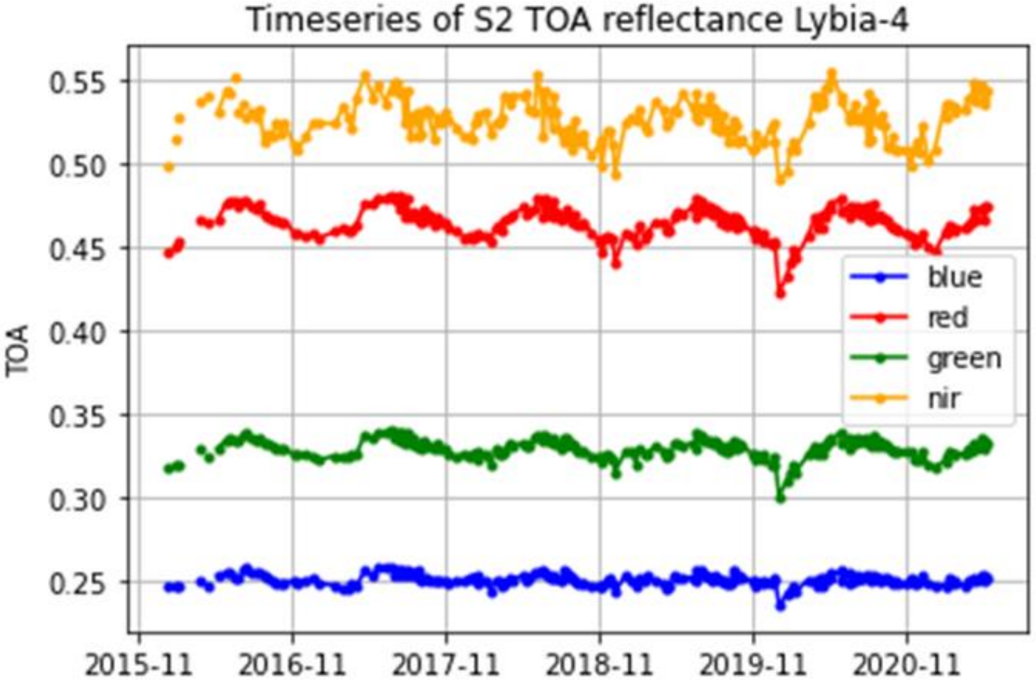
## Reading S2 products: define the function

*S2 data extraction over RoI*

```
def reading_s2 (path):

        b2=! find $path -name '*B02.jp2'
        b3=! find $path -name '*B03.jp2'
        b4=! find $path -name '*B04.jp2'
        b8=! find $path -name '*B08.jp2'

        with rasterio.open(b2[0]) as blue_band:
                #blue_band.crs (proprietà di rasterio ) accedo al CRS (coor. ref.system) del mio datas
                # con Proj (python interface di Proj) tras
                xlon,ylat=Proj(blue_band.crs)(lon,lat) # proietto lat lon in coord spaziali
                xmin=xlon-delta
                xmax=xlon+delta
                ymin=ylat-delta
                ymax=ylat+delta
                # e dalle coordinate spaziali passo alla posizione dei pixel da cui poi costruisco wi
                t = blue_band.transform
                col_min, row_min = ~t * (xmin, ymin)
                col_max, row_max = ~t * (xmax, ymax)

                window=Window.from_slices(rows=(floor(row_max), ceil(row_min)),
                                          cols=(floor(col_min), ceil(col_max)))
                blue_sub=blue_band.read(1, window=window)
        with rasterio.open(b3[0]) as green_band:
                green_sub=green_band.read(1, window=window)
        with rasterio.open(b4[0]) as red_band:
                red_sub=red_band.read(1, window=window)
        with rasterio.open(b8[0]) as nir_band:
                nir_sub=nir_band.read(1, window=window)
        blue=blue_sub/10000
        green=green_sub/10000
        red=red_sub/10000
        nir=nir_sub/10000
```

- Data extraction (S2 L1C) over 1kmx1km RoI
- Timeseries of TOA reflectances for the entire period of the mission
- Trend analysis of imagery acquired over a CEOS desert pseudo-invariant calibration site



# Outcomes (III) – Ready-to-use working environment

IDEAS-QA4EO

- The working environment is easily accessed over the internet:
  - **SSH (Secure Shell Protocol)**
  - **JupyterLab** (web access)
  - **X2Go client** (remote desktop access)
- All main **libraries for Cal Val activities** already installed, together with a **set of Notebooks** developed and successfully consolidated within the pilot activities
- Short time to access the **operating environment** and EO data archive

## Users activated on the platform:

| | |
|---|---|
| *Constant Mazeran (Solvo)* | Sentinel 3: OLCI data processing using the **Bright Pixel Correction algorithm -** marine reflectance validation over AERONET_OC sites |
| *Erminia De Grandis (Serco)* | Sentinel 2: L1C and L2A products assessment - **AOT and WV direct validation** with AERONET in-situ measurements - SR reference dataset built using **6S radiative transfer model** |
| *Gabriele Brizzi (Serco)* | *Sentinel 5p: L2* **products validation** against a subset of **in-situ measurements** AERONET and PANDONIA |
| *Daniele Di Erasmo (Serco)*<br>*Laura Fioretti (Aresys)* | IPF JERS2.07 and PALSAR 4.20 **processors** integration on the platform and **validation process** successfully completed |
| *Georgia Doxani (Serco)* | ACIX III Land: make use of the code developed during the S2 activities to implement the procedure to build the PRISMA **Surface Reflectance reference dataset** |
| *Alessandro Di Bella (Serco)* | Cryosat-2: analysis and processing of **Cryosat-2 data archive** |
| *David Fuertes (Grasp Sas)*<br>*J.C.Antuña Sanchez (Grasp Sas)* | GRASP **algorithm implementation and execution** - test the environment and investigate the possibilities to exploit this platform for their activities |

# Outcomes (IV) – Interoperability

IDEAS-QA4EO

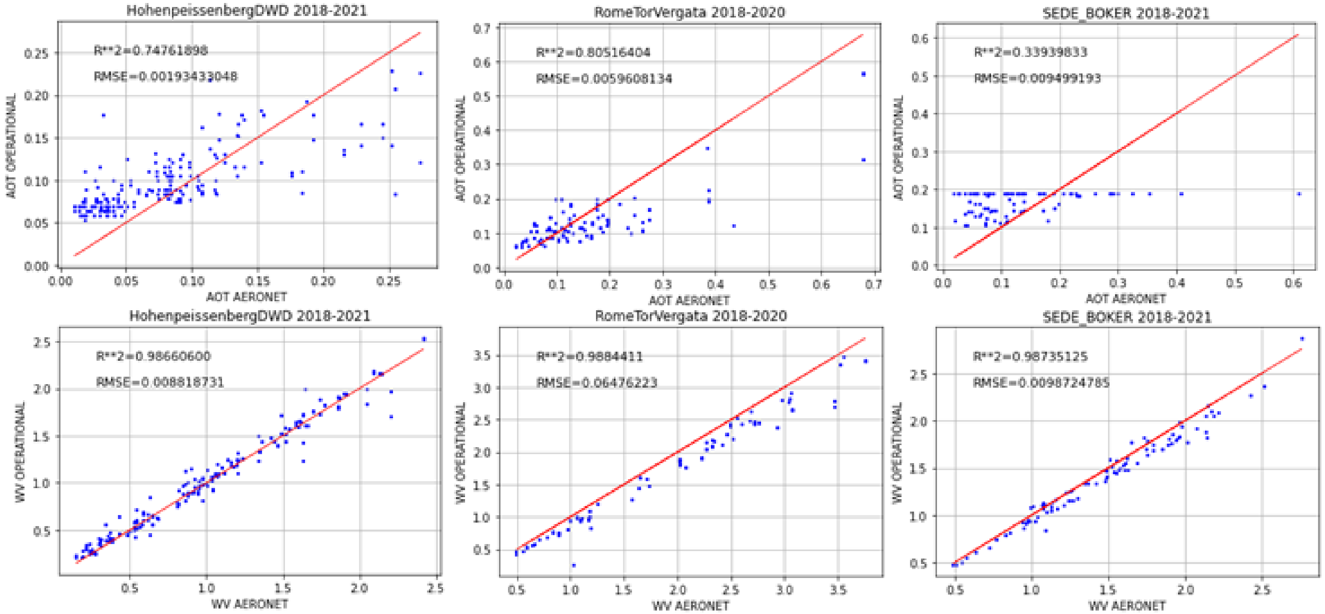
**Public knowledge library**: set of consolidated notebooks already employed by users working on the platform

## 1. Match-ups extraction with AERONET data

@INPUTS
- aeronet file level 2.0
- satellite acquisition time
- temporal window where to average AERONET measurements

@OUTPUTS
AERONET measurements averaged over defined temporal window centered at the satellite acquisition time

→ Sentinel-2 AOT/WV direct validation and inputs to 6SRTM
Sentinel-3 marine reflectance validation over AERONET-OC
Sentinel-5p Level2 products validation



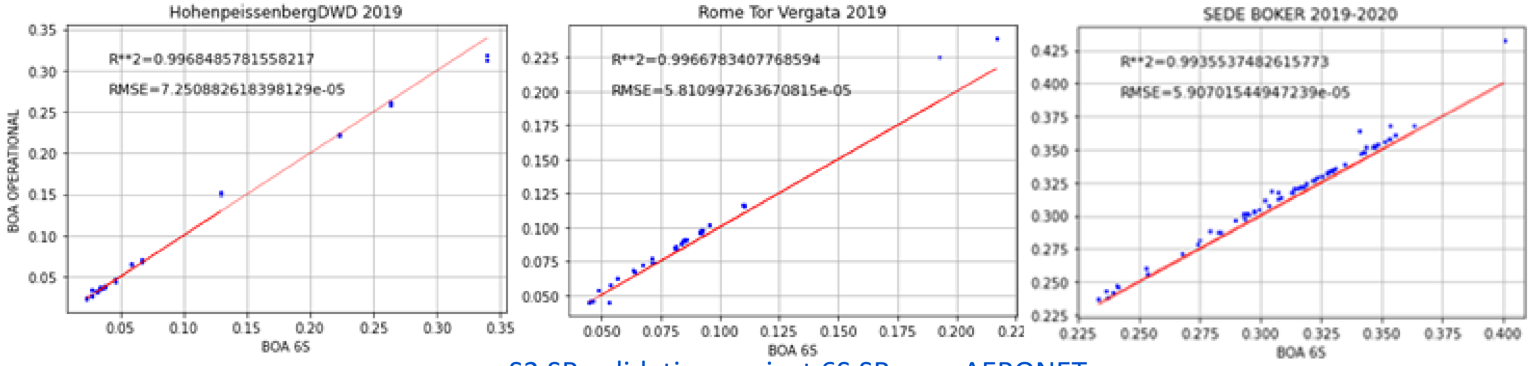

S2 AOT direct validation

S2 WV direct validation

## 2. 6SV Radiative Transfer Code

6S simulations using input AERONET state variables and TOA match-ups with AERONET data to build **synthetic surface reflectance**. The 6S object is built using the atmospheric and geometric information such as atmospheric profile, aerosol profile and observation geometry extracted by collocated data

→ Sentinel-2 L2A products validation
ACIX III – SR reference dataset for PRISMA products validation




S2 SR validation against 6S SR over AERONET

## 3. JERS2.07 and PALSAR 4.20 IPFs validation

The two processors have been implemented as testbed on the GBOX
The **SW verification process** within the QA4EO Facilities and Tools frame has been enhanced allowing **interoperability among different teams** in all steps of IPFs validation activity: SW installation, execution, verification and outputs validation

→ IPF validation process performed in the same environment: beneficial interoperability between the SW verification team and the Quality Control team

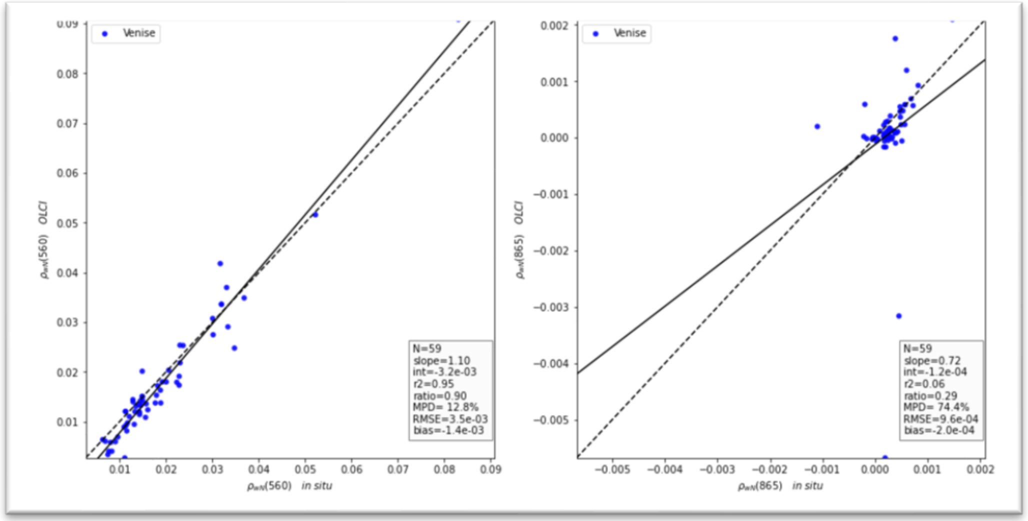
# Outcomes (V) – Processors and Algorithms integration

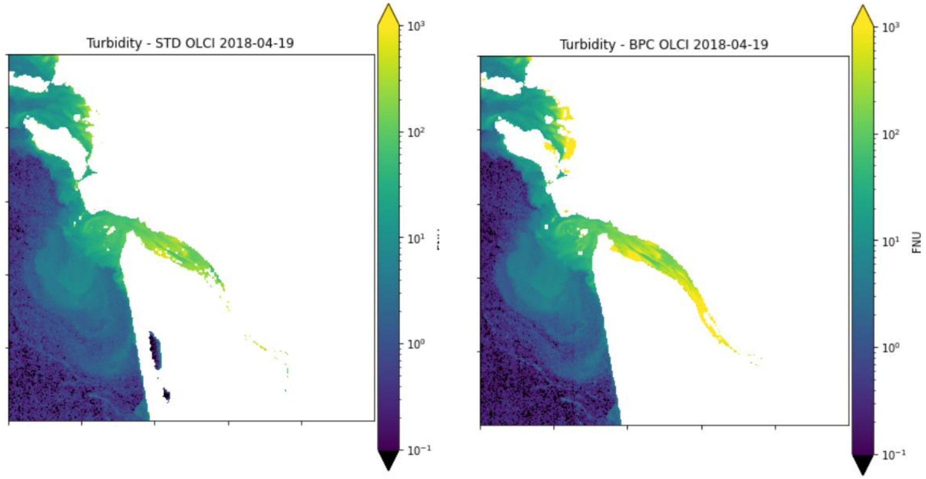
IDEAS-QA4EO

## Bright Pixel Correction algorithm

In-situ radiometric measurements of coast areas and match-ups with S3 OLCI Level-1 and Level-2 products

**BPC processor** (code of **Solvo** developed for EUMETSAT) to produce marine reflectance in the near-infrared – output validation with scatter plots and statistics

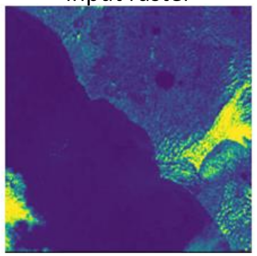



**Turbidity algorithm** (**Dogliotti** et al. 2015) and plot maps. Comparison exercise between operational Level 2 and BPC processor

## Sentinel 2 Masking Procedure

input raster



QI vector files

INPUTS

polygons in shape file

| | gml_id | maskType | geometry |
|---|---|---|---|
| 0 | OPAQUE.0 | OPAQUE | POLYGON ((234720.000 4700040.000, 235140.000 4... |
| 1 | OPAQUE.1 | OPAQUE | POLYGON ((239880.000 4699860.000, 240060.000 4... |
| 2 | OPAQUE.2 | OPAQUE | POLYGON ((244440.000 4700040.000, 244620.000 4... |
| 3 | OPAQUE.3 | OPAQUE | POLYGON ((224880.000 4699260.000, 225060.000 4... |
| 4 | OPAQUE.4 | OPAQUE | POLYGON ((223020.000 4699620.000, 223260.000 4... |
| ... | ... | ... | ... |
| 233 | OPAQUE.233 | OPAQUE | POLYGON ((283740.000 4603800.000, 283980.000 4... |
| 234 | OPAQUE.234 | OPAQUE | POLYGON ((298680.000 4602840.000, 298860.000 4... |
| 235 | OPAQUE.235 | OPAQUE | POLYGON ((283800.000 4600980.000, 283980.000 4... |
| 236 | OPAQUE.236 | OPAQUE | POLYGON ((286920.000 4595160.000, 287100.000 4... |
| 237 | OPAQUE.237 | OPAQUE | POLYGON ((288180.000 4593120.000, 288360.000 4... |

binary mask

masking

raster 1Kmx1Km

binary mask

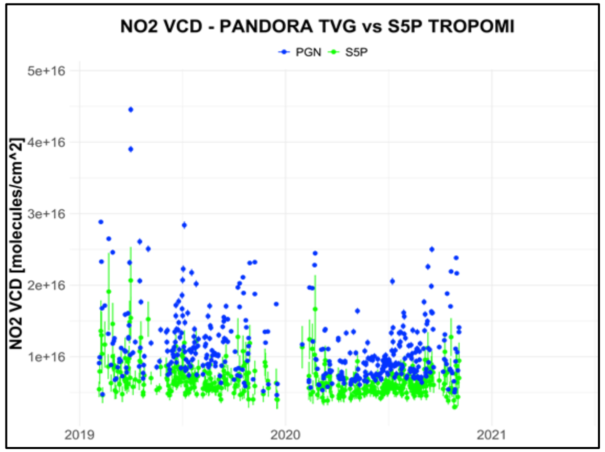
Clear, good data and not saturated pixels are selected

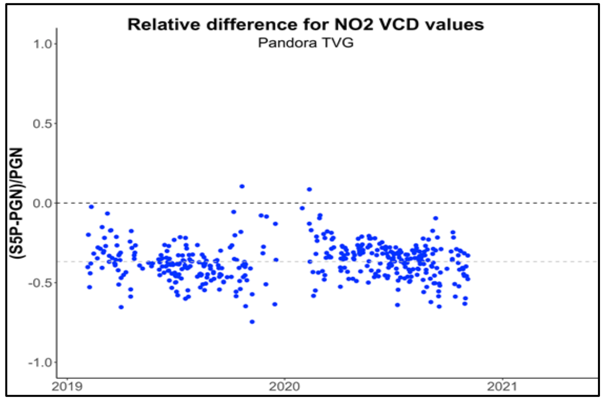
→ large set of data masked in a short time

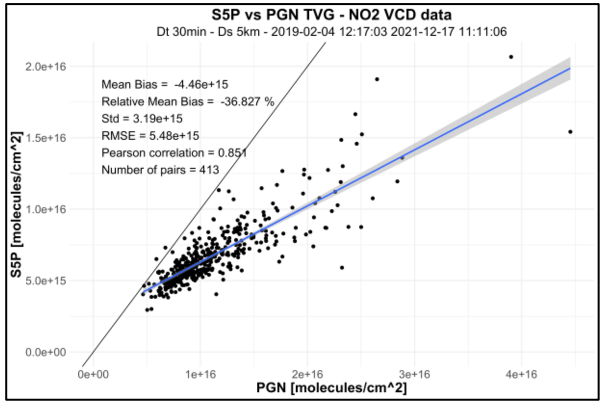
## GRASP algorithm

The GRASP team started testing the environment and investigating the data availability. They compiled the GRASP codes to gather feedback on the usage of the platform and on the advantages of a **potential parallel processing** or **multi-mission synergy processing**

# Outcomes (VI) – External datasets for products validation

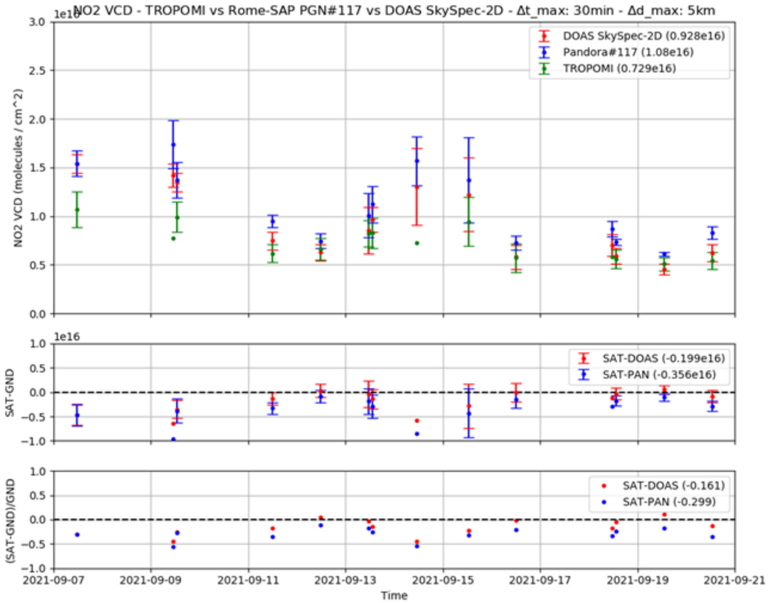


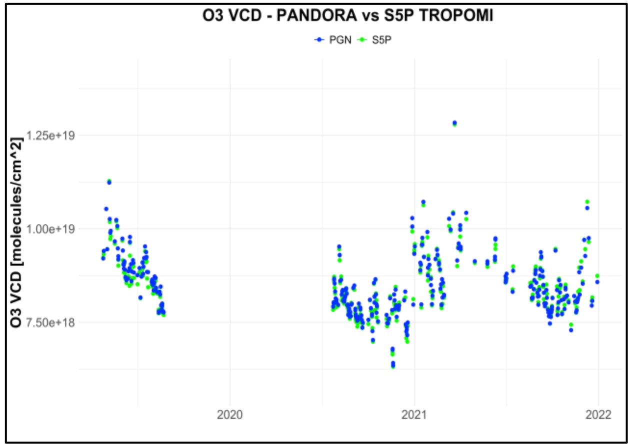

PGN#115

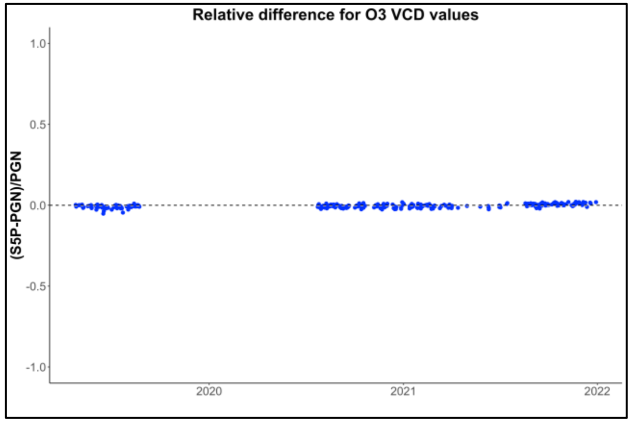
## S5P TROPOMI vs Pandora $NO_2$ and $O_3$ VCDs

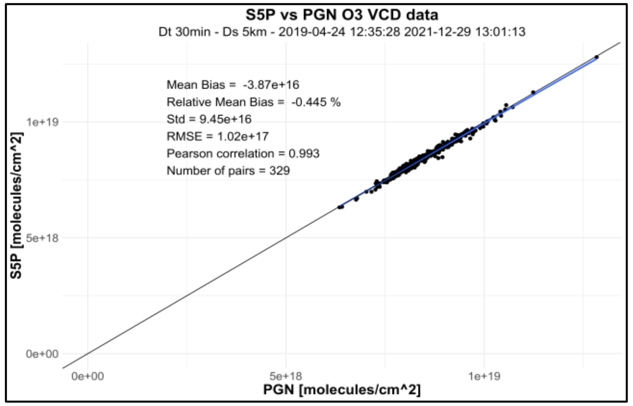



Intercomparison campaign BAQUNIN supersite
S5P v2.2 PGN V1.7

- Consolidate the validation procedure
- Extend the analysis to other PANDONIA stations including additional TROPOMI products
- Include AERONET network




PGN#117

# Conclusions and way forward

- March 2022 – end of demonstration phase (pilot):
  - The **Earth Console has proved** to meet the main requirements and to be a great solution in supporting the algorithms development and promoting a productive interoperability among the users
  - Despite the basic configuration of the Virtual Machine (multi-user environment) **we have clearly identified the advantages** provided by this environment that would enhance and support the activities within our team QA4EO also looking at the **active collaboration with the Progressive Systems team**
  - Considering all feedbacks and the potential future works, we have defined the best operational solution by **balancing user and system requirements with the associated cost model**

- April 2022 – start of the operational project

**Operational configuration**: Earth Console – GBOX IDE Type C
16 CPU - 64 GB RAM – 256 GB SSD
6TB STORAGE – Multi Users – CREODIAS archive

→ HW Resources increasing

## Next steps

- Integration in **Earth Console P-PRO** via I-APP of the applications developed in the use cases to enable scalable and **parallel executions**. Possibility to invocate P-PRO from the G-BOX
- **Data availability assessment**
- **Enrich** the collections of available software and tools and **ready-to-use notebooks**
- **Increase awareness** about the project, with dedicated promotional and outreach activities, dedicated web page with tutorials, success stories, results

→ LPS 2022 Bonn 23-27 May 2022 Session C5.03 '*Open Source, Science, toolboxes and Jupyter technologies in EO*'

# Evolutions

The EarthConsole® is a **user-centered platform** where the development of new features is driven by user needs

- **Containerize processors** in order to abstract their execution from the underlying computing infrastructure. It will be possible to run processors containers customizing input data and parameters, and to **monitor the container status and resources usage**
- Use of **profiling tools** to support processors performance monitoring and optimization
- Support parallel processing (P-PRO) invocation from the G-BOX as well as retrieval of the results, via dedicated APIs (OGC compliant)
- **Setup of Virtual Lab as a collaborative space** for sharing code, data, results, documents and information among different users: the all-in-one place to access the EarthConsole® services

For example: Altimetry Virtual Lab (already operational) https://earthconsole.eu/virtual-labs/

# QA4EO Interoperability Infrastructure for Cal Val Activities

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | S | F | E | L | G | M | T | S | B | I | Z | N |
| P | N | G | E | R | E | Q | R | T | I | B | M | S |
| X | H | T | E | W | L | C | R | A | K | S | L | I |
| S | N | R | H | B | O | Z | U | F | B | D | K | R |
| L | Z | G | P | Y | P | E | S | Q | V | W | T | Y |

## Q & A

*IDEAS-QA4EO Cal/Val Workshop#3 - March 31, 2022 - ESA/ESRIN*