



# **The European Commission's science and knowledge service**

## **Joint Research Centre**

# Copernicus DIAS for CAP “checks by monitoring”

Guido Lemoine ([guido.lemoine@ec.europa.eu](mailto:guido.lemoine@ec.europa.eu)), Konstantinos Anastasiakis

# Copernicus DIAS for CAP CbM

DIAS = Data and Information Access Systems, the Copernicus programme effort to facilitate cloud based data analytics for value adding to Copernicus data

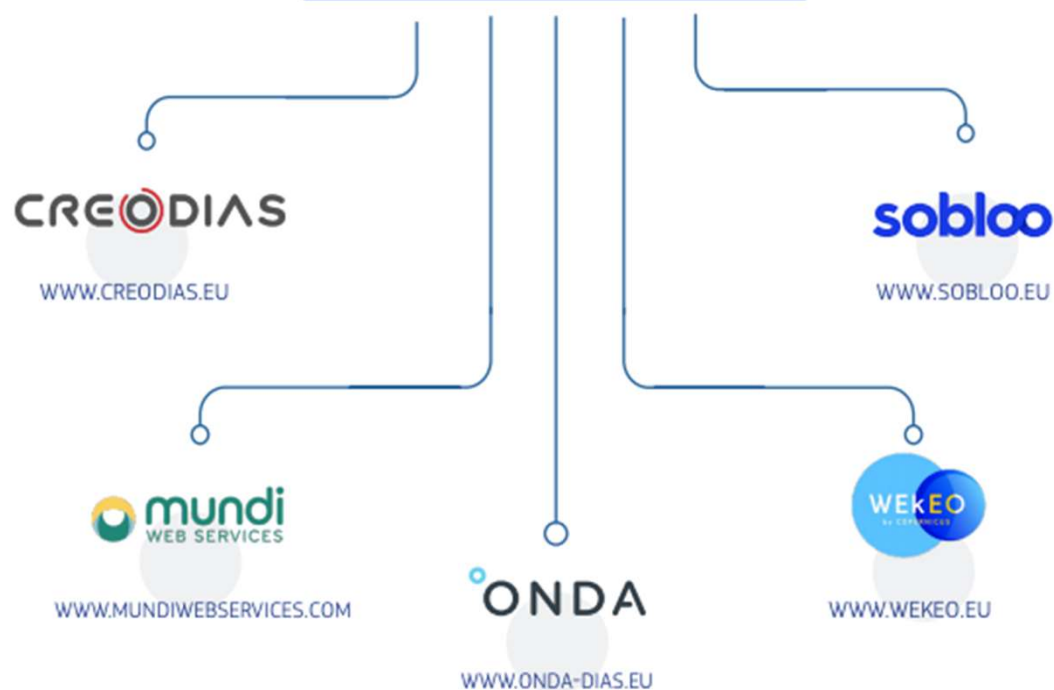
DIAS initiated in late 2017, operational readiness in June 2018

CAP Checks by Monitoring (CbM) is potential application domain, where coordinated value adding on DIAS platform may bring operational benefits

Basis for decision by DG GROW to support DG AGRI in “DIAS onboarding” of Paying Agencies which opted for CbM by 1 January 2018 (ES, IT, BE (F), DK, MT)

Through ESA work orders, with the technical supervision of JRC D.5

## THE DIAS & WHERE TO REACH THEM



## 5x DIAS

WEkEO not relevant (no S1 and S2, for atmosphere & oceans)

# What's a DIAS?

The full Sentinel data archive (both ESA generated data, Copernicus Services outputs)

Closely coupled with cloud computing infrastructure for on-demand processing

A development environment for “third parties” to generate higher level outputs and services, for commercial exploitation

Various degrees of support, access to additional data, shared business, etc.

Only self-sustained DIAS will continue as fully commercial instances (> 2021)

# What's a DIAS (the tech view)?

**Object storage** giving access to [partial] copies of the ESA scihub archive

**An interface** to find what is available on the DIAS

**A platform** to select and configure cloud resources for processing

**Configurable processing** to run in “front end” or “backend”

**Documentation** [on line, off line] on the various components and technical issues

**Technical support** for IT infrastructure and (through partners) EO specific issues

# What's object storage?

Object storage is the preferred storage for immutable “Big Data” (PBs) storage

Write once, read often (e.g. YouTube, Google Earth Engine, AWS) as indexed blobs

Much simpler to manage and extend than file or block storage, much cheaper

But requires API to transfer to classical file system, handle as “normal” file

CREODIAS, MUNDI and SOBLOO all use S3 (AWS, GCS standard) object storage. ONDA uses ENS (OpenStack Swift).

# How to find relevant data?

DIAS stores Level 1C by default. For CbM they are required to generate Level 2. S1: **CARD-BS, CARD-COH6**, S2: **CARD-2A**

All goes in the object storage, including CARD products generated on the backend.

[CREODIAS](#) and MUNDI publish their data sets in their on-line catalog.

Catalog can be searched via web interface or OpenSearch compatible queries.

SOBLOO provides a postgis data base.

Query provides S3 endpoint. Transfer with s3cmd or python (boto3)



## SEARCH CRITERIA

Search phrase, e.g. winter in Quebec

Product identifier or path

observed  2019-04-01  2019-04-03

published  YYYY-MM-DD  YYYY-MM-DD

position latitude longitude

cloud cover 0-100 %

show only local products ☐

collection:

Sentinel-1

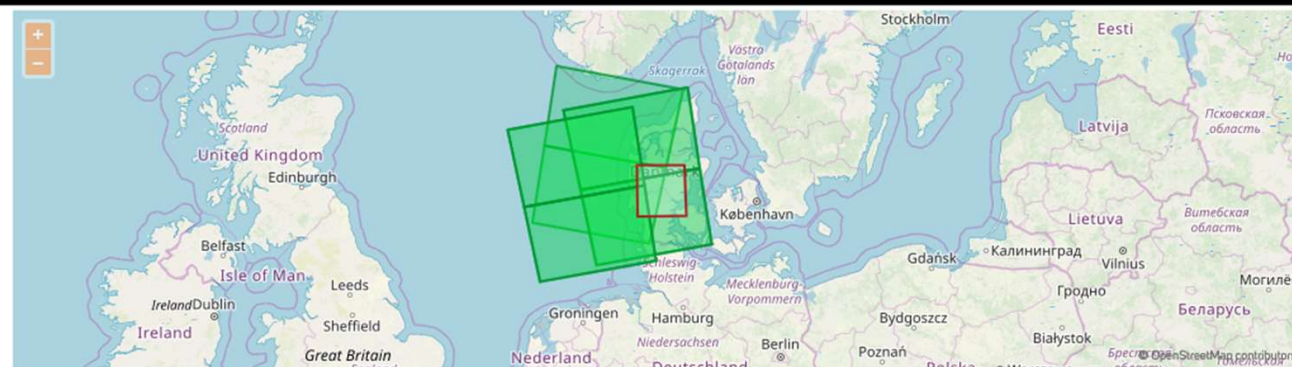
product type: CARD-BS

<https://finder.creodias.eu/resto/api/collections/Sentinel-1>

Polygon Selection  Upload Polygon 

Point Selection  Clear All 

Search



## search results

Title	Observation date	Publication date	Cloud %	File size
S1B_IW_GRDH_1SDV_20190403T054818_20190403T054843_015638_010532_3B6...	2019-04-03 05:48:18.861	2019-04-03 13:23:10.049035	-	4.88GB
S1B_IW_GRDH_1SDV_20190403T054753_20190403T054818_015638_010532_5D7F...	2019-04-03 05:47:53.862	2019-04-03 13:24:04.519028	-	4.87GB
S1A_IW_GRDH_1SDV_20190402T170949_20190402T171014_026614_02FC31_8F0E...	2019-04-02 17:09:49.26	2019-04-03 11:58:52.551964	-	4.88GB
S1A_IW_GRDH_1SDV_20190402T170949_20190402T171014_026614_02FC31_D7B8...	2019-04-02 17:09:49.26	2019-04-03 11:58:41.276798	-	4.88GB
S1A_IW_GRDH_1SDV_20190402T170924_20190402T170949_026614_02FC31_1F29...	2019-04-02 17:09:24.26	2019-04-03 11:59:14.938884	-	4.72GB
S1A_IW_GRDH_1SDV_20190402T170924_20190402T170949_026614_02FC31_6ED2...	2019-04-02 17:09:24.26	2019-04-03 11:59:24.561382	-	4.72GB
S1B_IW_GRDH_1SDV_20190401T171715_20190401T171740_015616_01D477_3A01_CAR...	2019-04-01 17:17:15.535	2019-04-01 22:08:02.338193	-	5.03GB
S1B_IW_GRDH_1SDV_20190401T171715_20190401T171740_015616_01D477_3A08_CA...	2019-04-01 17:17:15.535	2019-04-02 04:07:37.689584	-	5.03GB
S1B_IW_GRDH_1SDV_20190401T171650_20190401T171715_015616_01D477_6F13_CA...	2019-04-01 17:16:50.536	2019-04-02 04:15:29.077721	-	4.95GB
S1B_IW_GRDH_1SDV_20190401T171650_20190401T171715_015616_01D477_C7C1_CA...	2019-04-01 17:16:50.535	2019-04-01 22:14:25.298537	-	4.95GB

showing 10 out of 10 total result(s)

copy all as paths

copy all as urls

add all to cart

remove all from cart

# Defining cloud computing resources

The DIAS user can set up and configure virtual (or dedicated) machines to perform computing [on DIAS data].

Via web interface (all, except ONDA) or [OpenStack API \(CREODIAS\)](#), or with technical support of the DIAS provider (all)

On DIAS platform infrastructure (e.g. Orange, OVH, Telekom)

Configurable CPU, RAM, disks, object storage, network setup, etc.

All provide configured linux images with pre-installed required software tools (see HLR documents). All support docker (swarm).

User can access machines via ssh.



## Cloud Server Console

Dashboard

Elastic Cloud Server

Cloud Server Backup Service

Bare Metal Server

Elastic Volume Service

Dedicated Storage Service

Volume Backup Service

Image Management Service

Auto Scaling

Key Pair

ECS Group

## Elastic Cloud Server

+ Create ECS

You can create 91 more ECSs. The ECSs can use up to 703 vCPUs and 1,215 GB of memory. [Quota details](#)

Start Stop Restart Delete

All statuses

Name



Search by Tag



<input type="checkbox"/>	Name/ID	AZ	Status	Specifications/Image	IP Address	Operation
<input type="checkbox"/>	sobloo-vm-0005 d40e969d-36e8-45a4-8f79-...	eu-west-0b	Running	16 vCPUs   64 GB   s3.4... OBS Ubuntu 18.04	192.168.30.105 (Private IP)	Remote Login...
<input type="checkbox"/>	sobloo-vm-0001 f4b5a44f-405b-4b68-9c29-...	eu-west-0b	Running	16 vCPUs   64 GB   s3.4... OBS Ubuntu 18.04	192.168.30.173 (Private IP)	Remote Login...
<input type="checkbox"/>	sobloo-vm-0004 a085f799-cd7b-4284-8e70-...	eu-west-0b	Running	16 vCPUs   64 GB   s3.4... OBS Ubuntu 18.04	192.168.30.229 (Private IP)	Remote Login...
<input type="checkbox"/>	sobloo-vm-0002 9c60687c-c4f4-4fc4-8d39-...	eu-west-0b	Running	16 vCPUs   64 GB   s3.4... OBS Ubuntu 18.04	192.168.30.129 (Private IP)	Remote Login...
<input type="checkbox"/>	sobloo-vm-0003 1d1c1500-2d0c-4065-a50e-...	eu-west-0b	Running	16 vCPUs   64 GB   s3.4... OBS Ubuntu 18.04	192.168.30.231 (Private IP)	Remote Login...
<input type="checkbox"/>	bastion-cce1 1cc6eda6-c771-4115-bc67-...	eu-west-0b	Running	1 vCPUs   1 GB   t2.micro OBS Ubuntu 18.04	90.84.243.7 (EIP) 100 Mbit/s 192.168.30.77 (Private IP)	Remote Login...
<input type="checkbox"/>	reserved-6vi6b 207afd0b-9980-48a9-ab93-...	eu-west-0a	Running	4 vCPUs   16 GB   s3.xla... cce-node-1.11.3-2019-0...	192.168.30.174 (Private IP)	Remote Login...
<input type="checkbox"/>	reserved-v3kzc 382dfc29-fe46-47ba-8e7f-5...	eu-west-0a	Running	4 vCPUs   16 GB   s3.xla... cce-node-1.11.3-2019-0...	192.168.30.238 (Private IP)	Remote Login...
<input type="checkbox"/>	desktop 1c8538c9-4678-4603-a14f-...	eu-west-0a	Running	8 vCPUs   32 GB   s3.2xl... OBS_Ubuntu_18.04	90.84.246.50 (EIP) 1,000 M... 192.168.42.96 (Private IP)	Remote Login...



# CfM work order phase 1

Check “technical readiness” of DIAS instances for initial CbM requirements

Generate CARD, both S1 geocoded backscattering and coherence, S2 atcorred, using standard software solutions (s1tbx, SEN2COR).

Provide initial compute environment, with scaling in mind

First impression on portability issues

For common pilot area (Denmark) in support to benchmarking (560 K parcels)





# Phase 2 JRC focus

On-boarding the Member States on assigned DIAS

Implement functionality in docker for deployment in docker swarm. Portability amongst (3) DIAS is then a minor issue.

Demonstrate full machine learning approach (port from GEE/tensorflow)

Performance testing, scaling, testing alternative data formats

Initial estimates for cost comparison

Phase 2 to be concluded in next weeks.

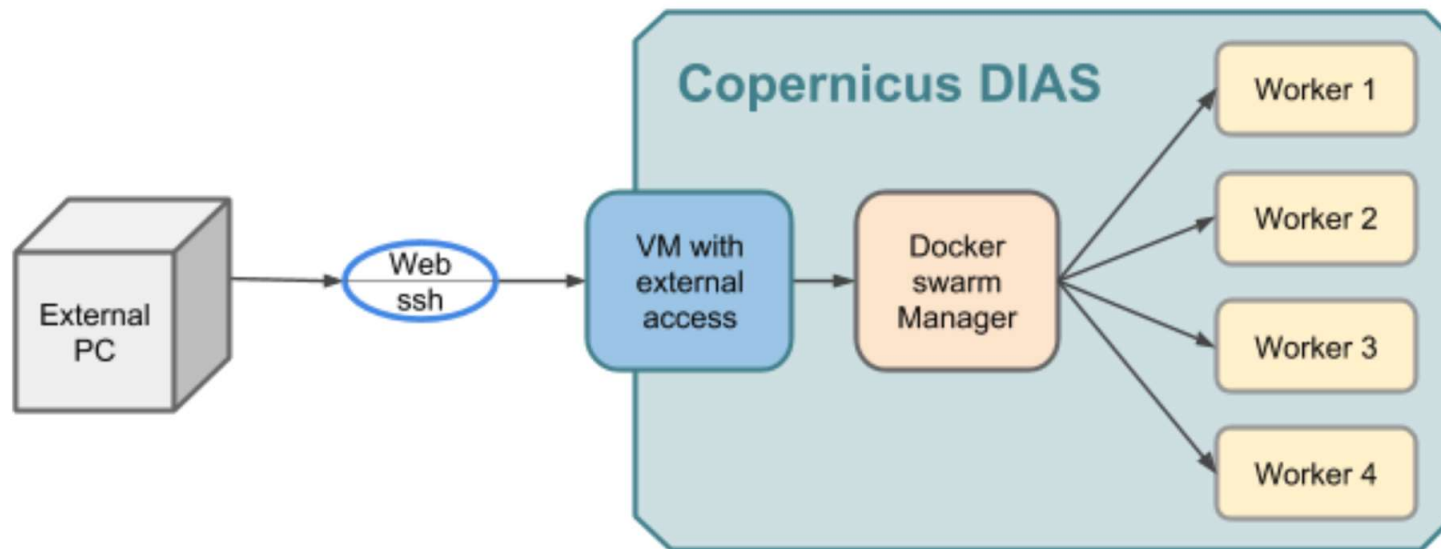
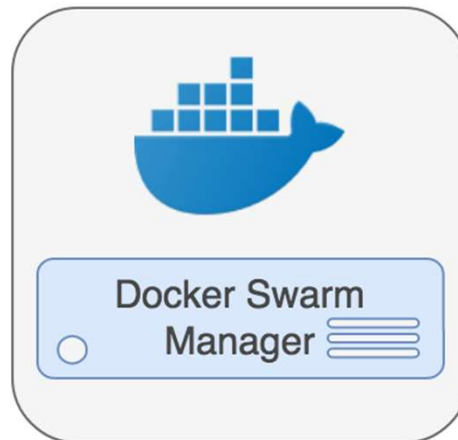
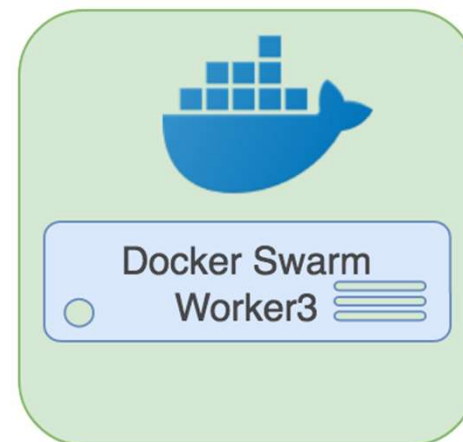
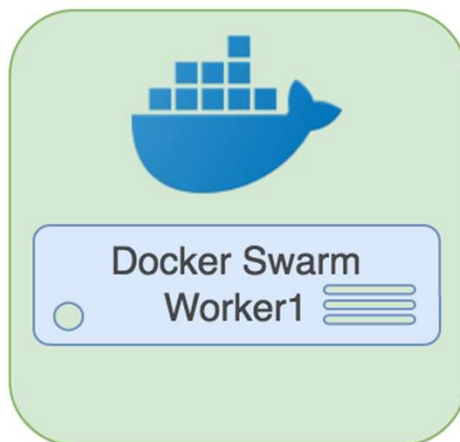


Illustration 4.1. Basic DIAS cloud architecture

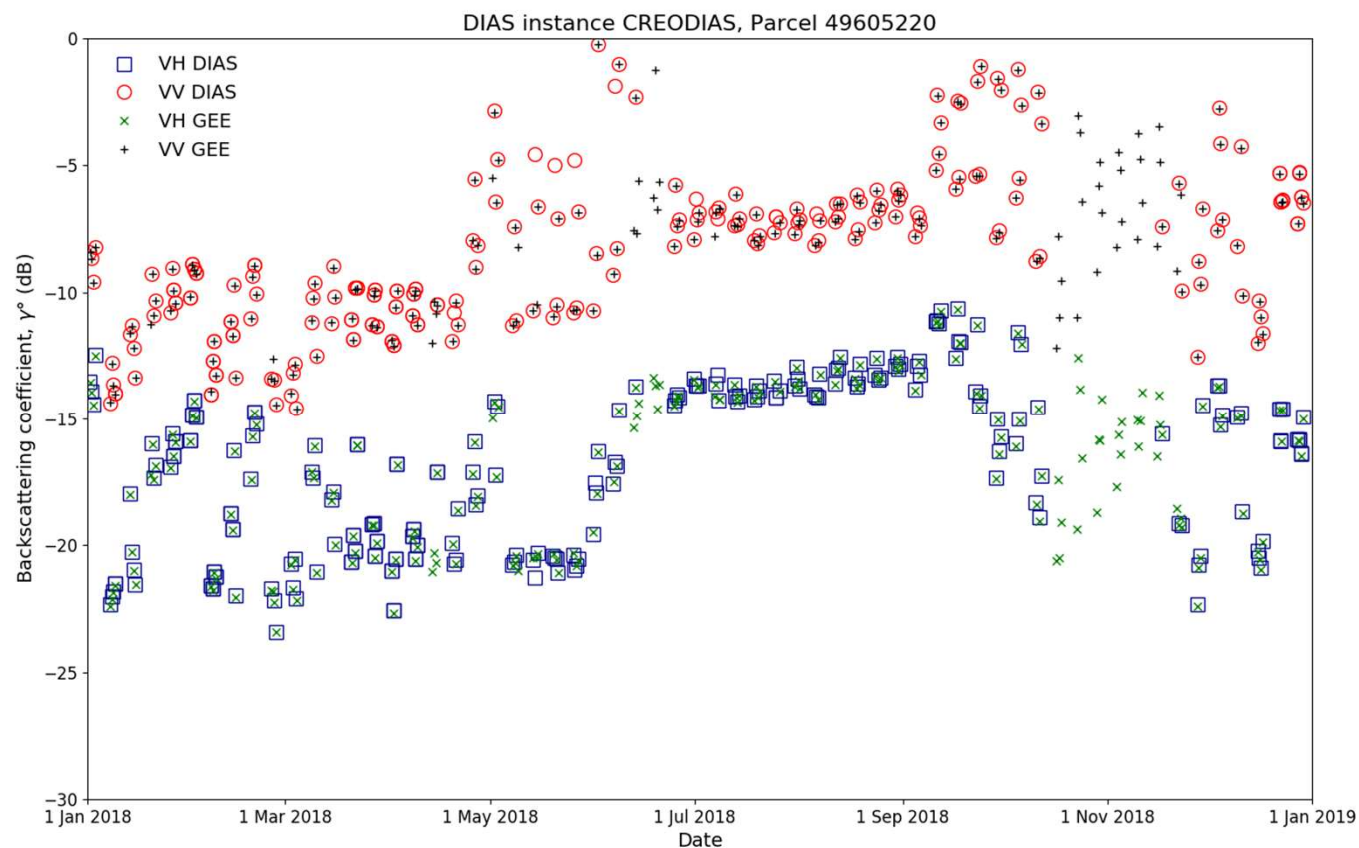
**Manager node hosts database that provides S3 endpoints and parcel vector and stores extracts. Uses postgresql/postgis image (mdillon/postgis).**



**Vector extracts run on worker nodes. Each worker has access to S3 store. Uses python3+rasterio+boto3 image. (glemoine62/dias\_py)**







CARD-BS per parcel

49605220 is a potato field

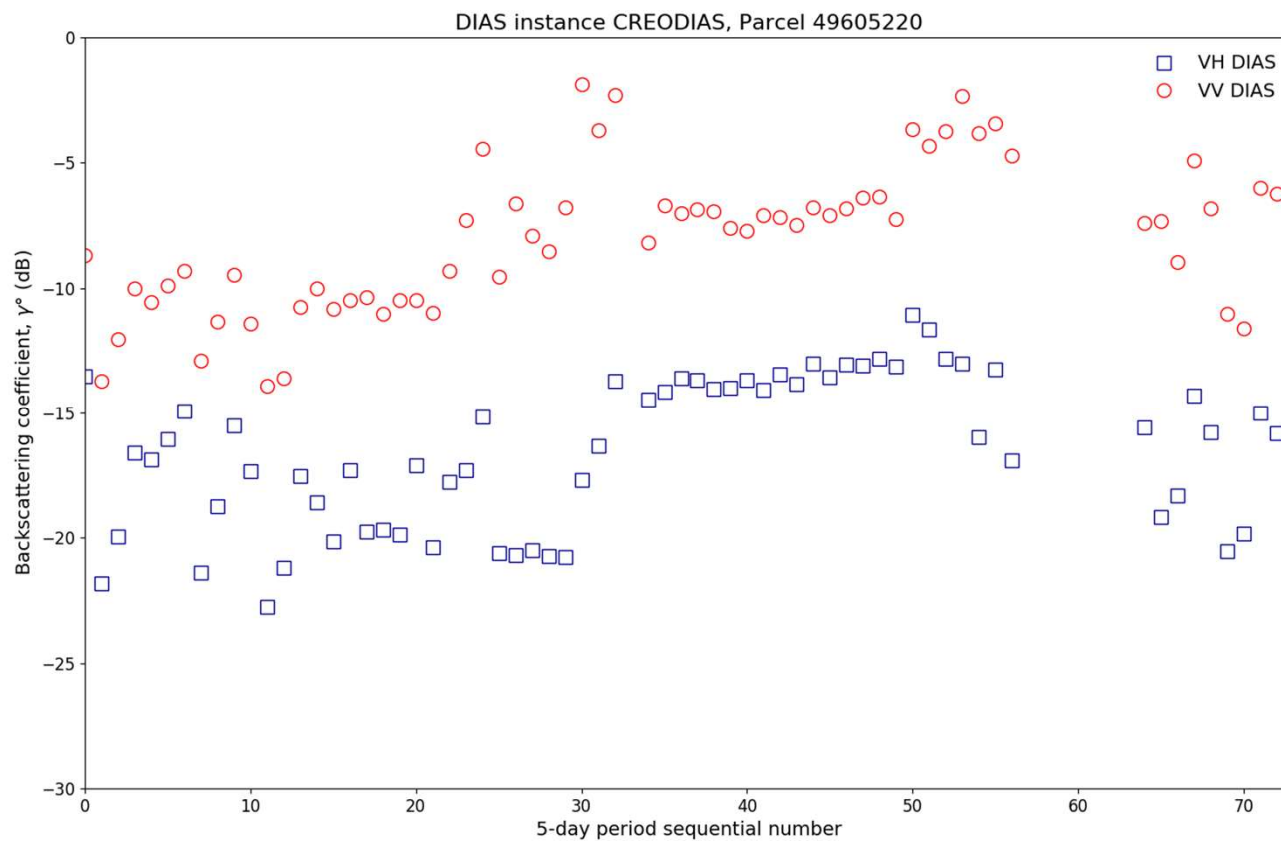
270+ million records for 2018, 2 days

Consistency compared to Google Earth Engine

Using same s1tbx graph

Same for other DIAS

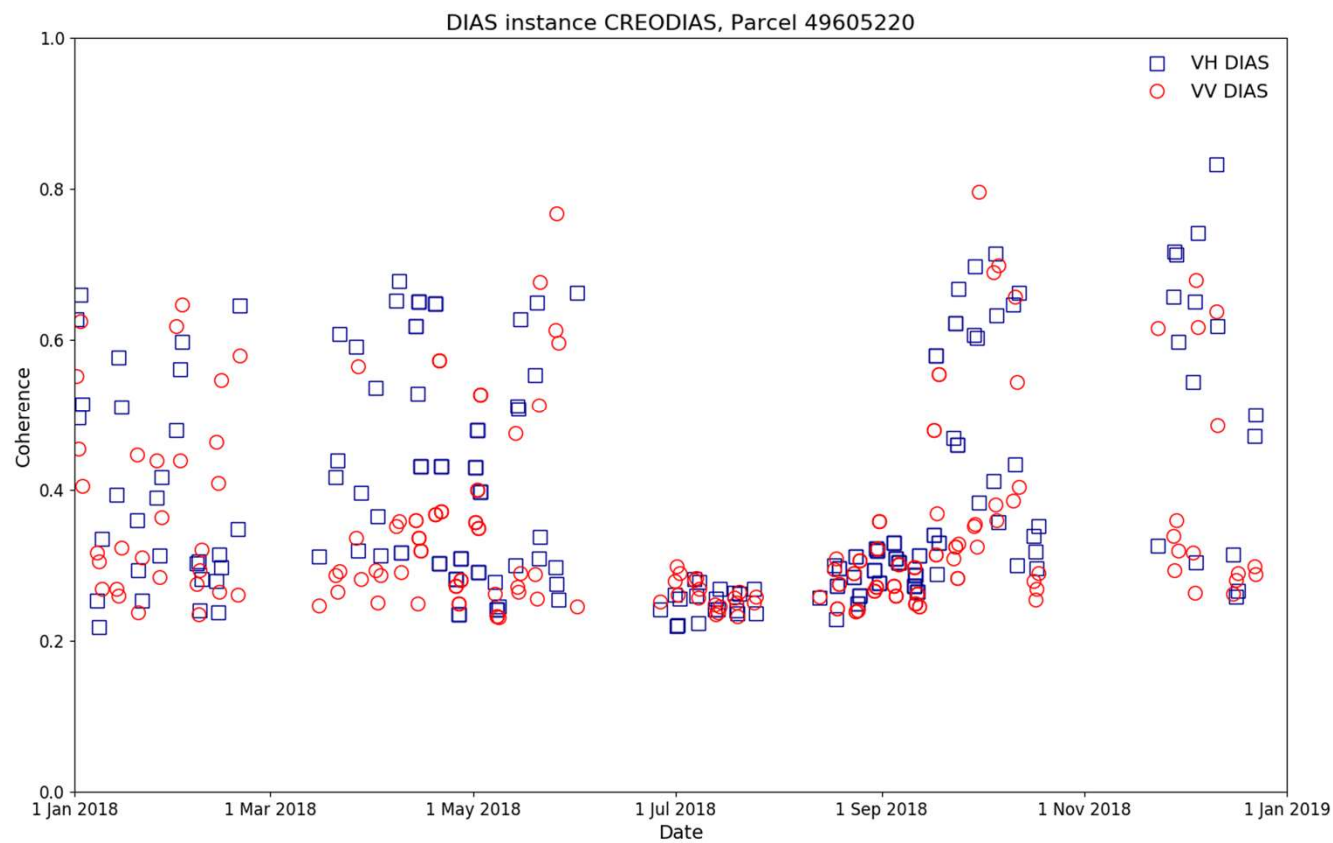
Some gaps in CREODIAS set



Averaged for 5 day periods

75+ million records for 2018

Input to machine learning



CARD-COH6 per  
parcel

6-day coherence

250+ million records  
for 2018

Not available  
anywhere else!

Some gaps due to  
failed S1 orbits

# Conclusions

DIAS are ready for CAP Checks by Monitoring!

DIAS is not Google Earth Engine. Core IT expertise in cloud computing, programming required.

DIAS may be used at increasing levels of complexity, e.g. (1) simple CARD downloads, (2) extracting reductions to (3) full data analytics

Level 2 and 3 are more interesting for shared method development

Only for Member States opting for CbM in 2019. But others can already benefit from experience gained.

JRC support for method testing and demonstration of common interest will continue in Phase 3.

# Next steps

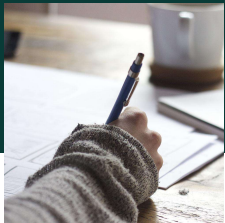
Further escalating technical solutions (kubernetes, machine learning)

S3 and fast sub-image access (Cloud Optimized GeoTIFF?) for image processing, interactivity

Further development of relevant PA routines (extract, reporting)

Integration and testing of open source SEN4CAP routines

(Phase 2 routines available on request, will be open sourced)



# Thanks!

## Any questions?

Contact [guido.lemoine@ec.europa.eu](mailto:guido.lemoine@ec.europa.eu)