The European Commission's science and knowledge service

Joint Research Centre

European Commission

Copernicus DIAS for CAP "checks by monitoring"

Guido Lemoine (guido.lemoine@ec.europa.eu), Konstantinos Anastasiakis

European Commission

Copernicus DIAS for CAP CbM

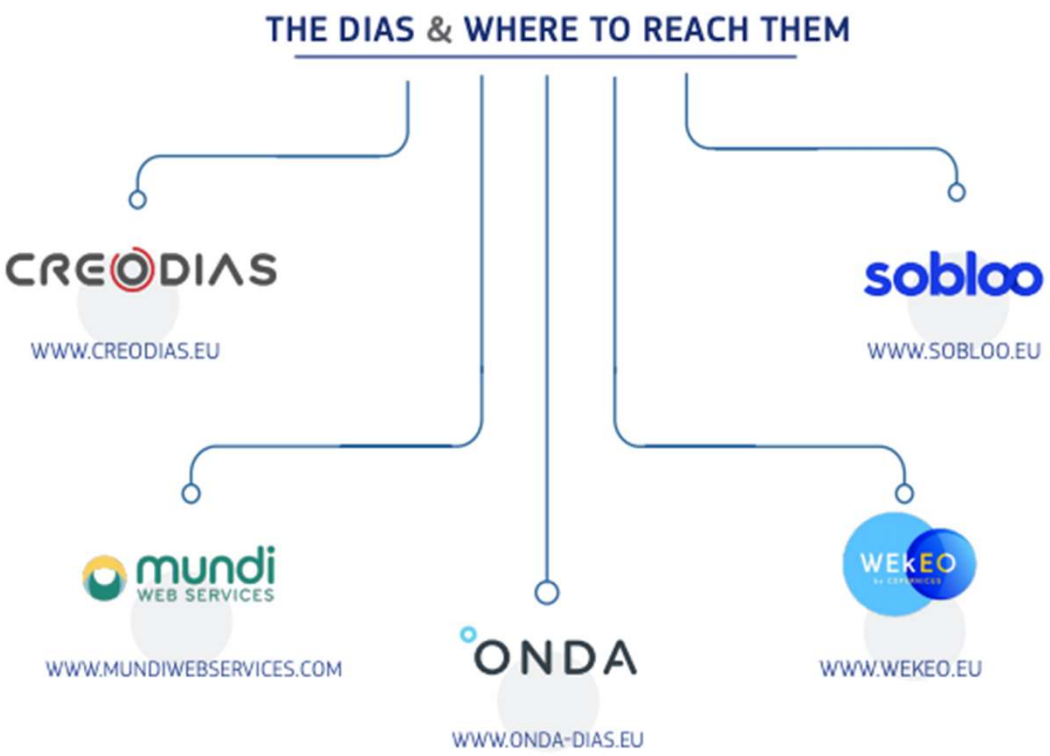
DIAS = Data and Information Access Systems, the Copernicus programme effort to facilitate cloud based data analytics for value adding to Copernicus data

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



5x DIAS

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

European Commission

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**

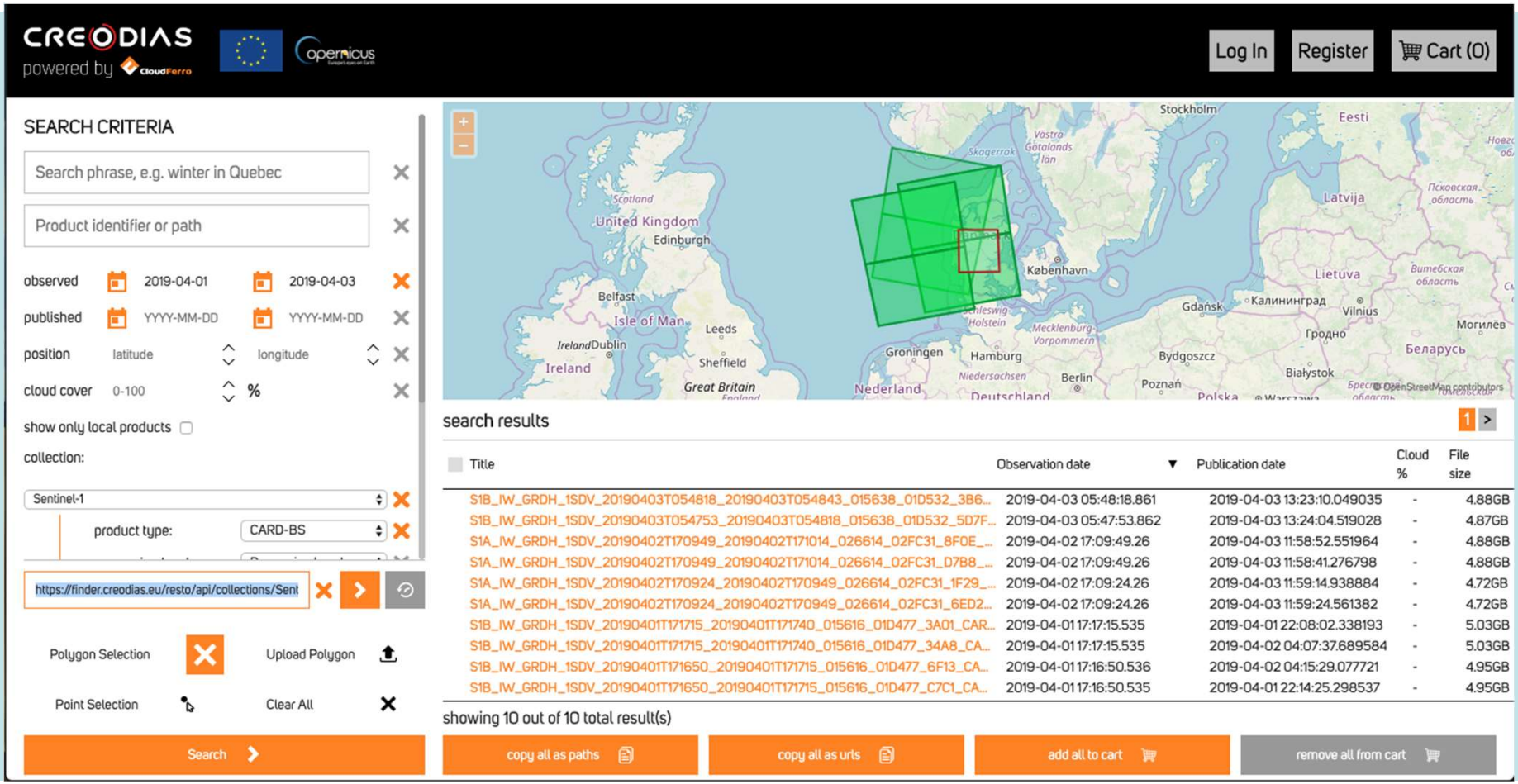
All goes in the object storage, including CARD products generated on the backend.

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)



European Commission

Defining cloud computing resources

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

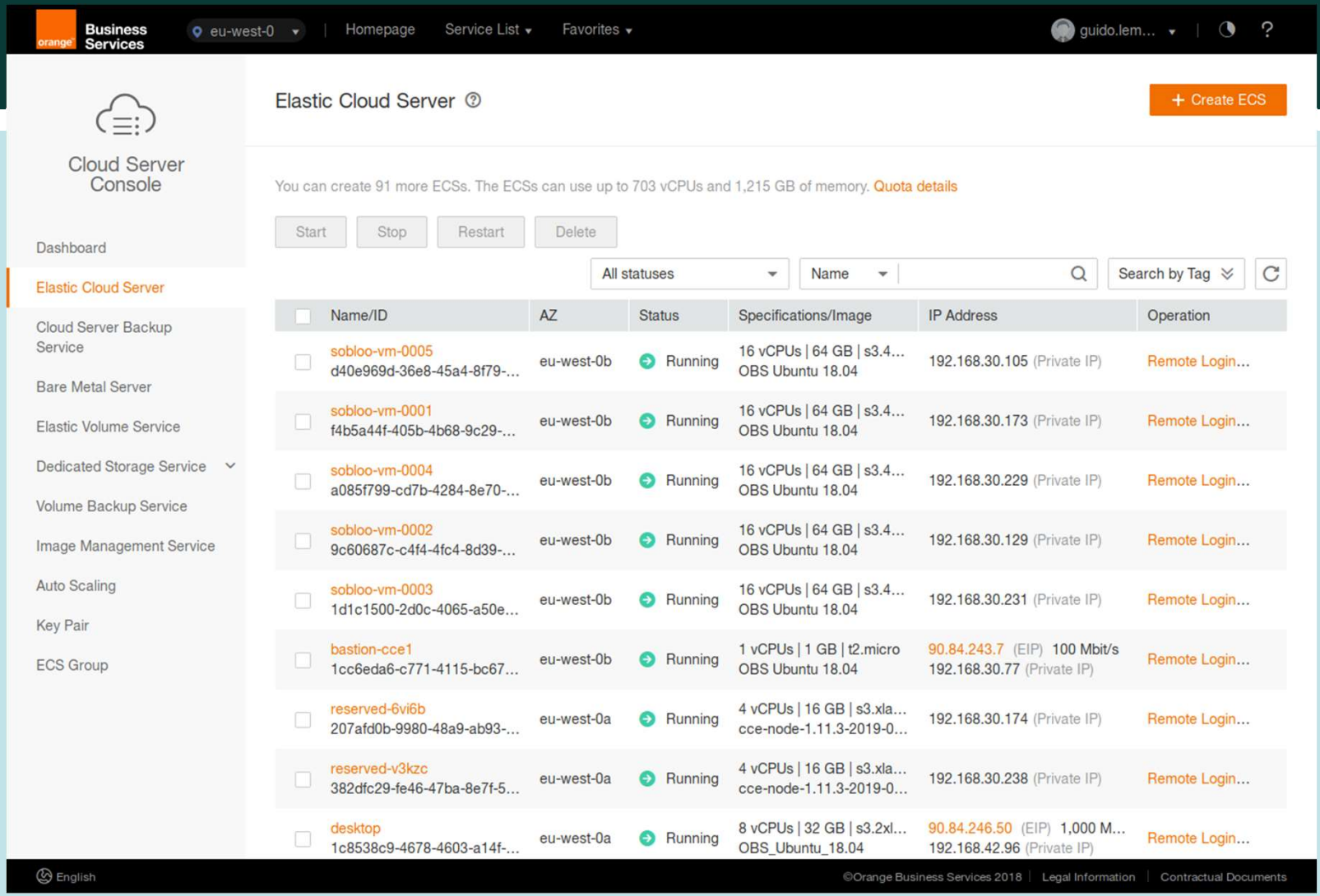
Via web interface (all, except ONDA) or OpenStack API (CREODIAS), or with technical support of the DIAS provider (all)

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.



European Commission

CfM work order phase 1

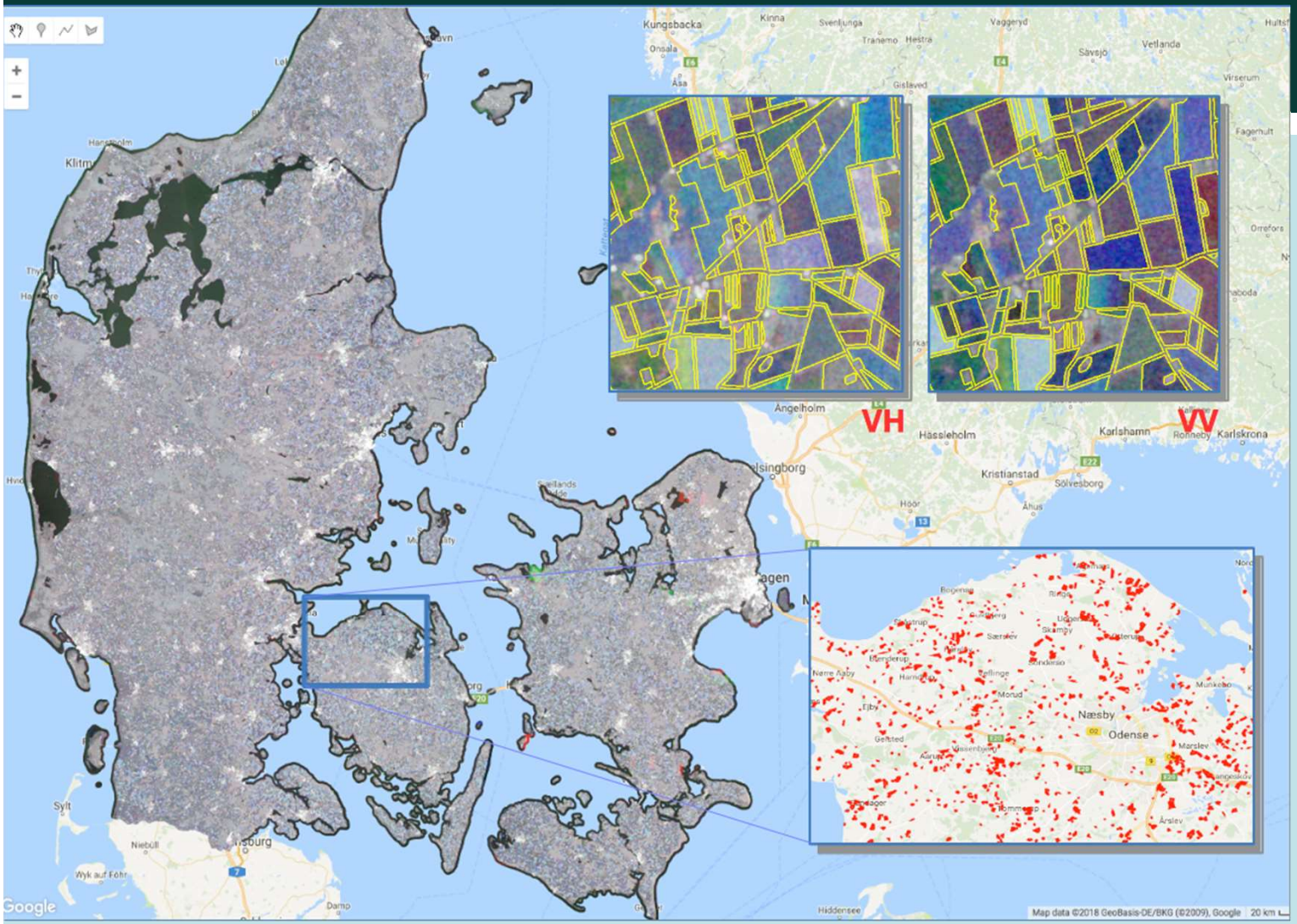
Check “technical readiness” of DIAS instances for initial CbM requirements

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)



S1 provides consistency (~200 images per year)
Machine learning applied to S1 time series
Splits 95-5 "outliers"
Follow up with additional imagery (S2) and other measures

European Commission

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.

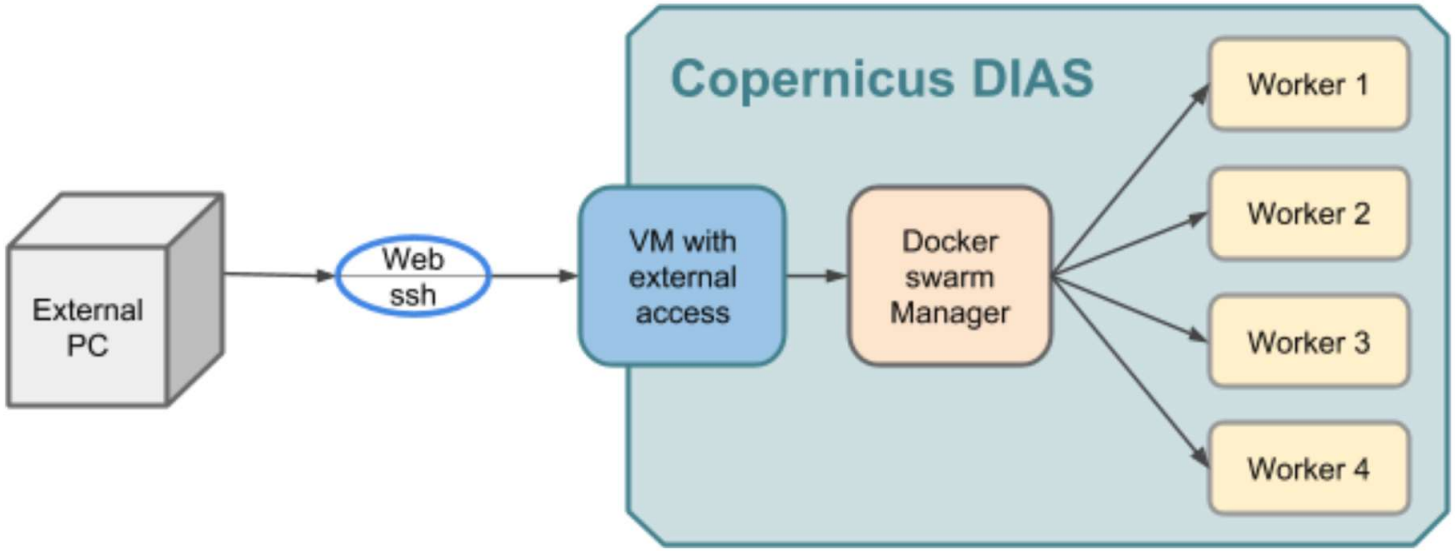
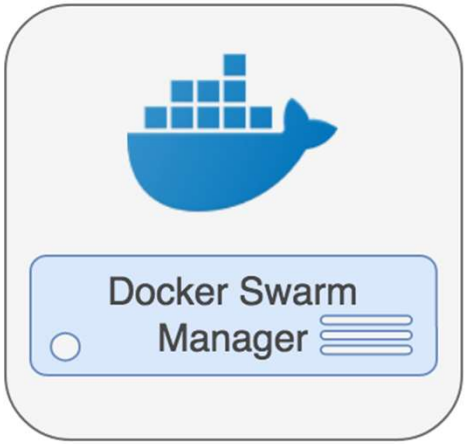
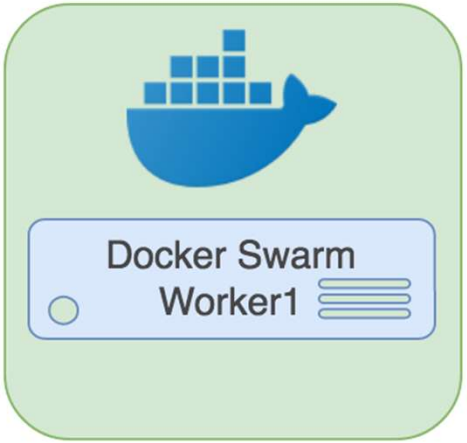


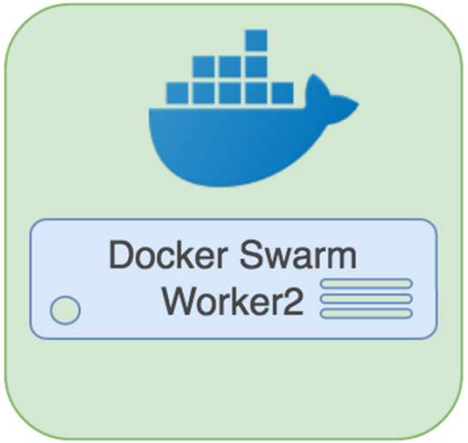
Illustration 4.1. Basic DIAS cloud architecture

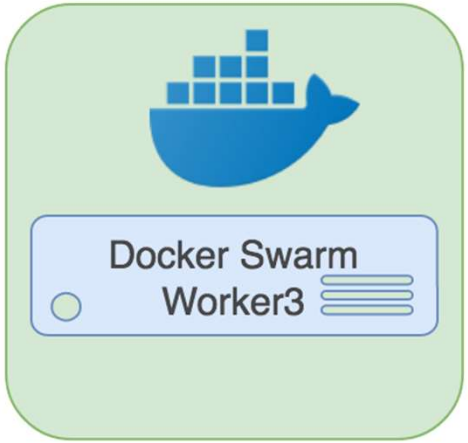
Manager node hosts database that provides S3 endpoints and parcel vector and stores extracts. Uses postgresql/postgis image (mdillon/postgis).

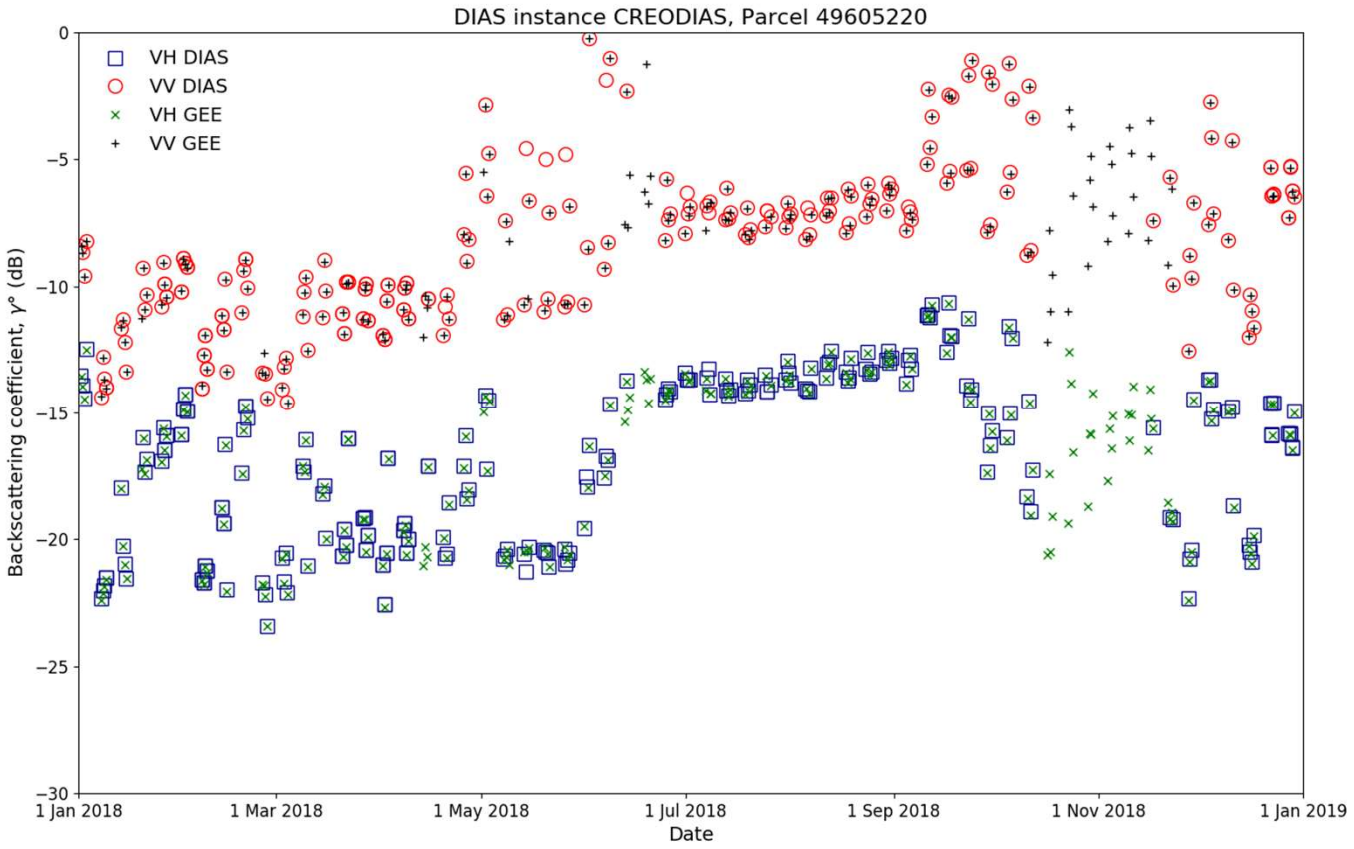


Vector extracts run on worker nodes. Each worker has access to S3 store. Uses python3+rasterio+boto3 image. (glemoine62/dias_py)









CARD-BS per parcel

49605220 is a potato field

270+ million records for 2018, 2 days

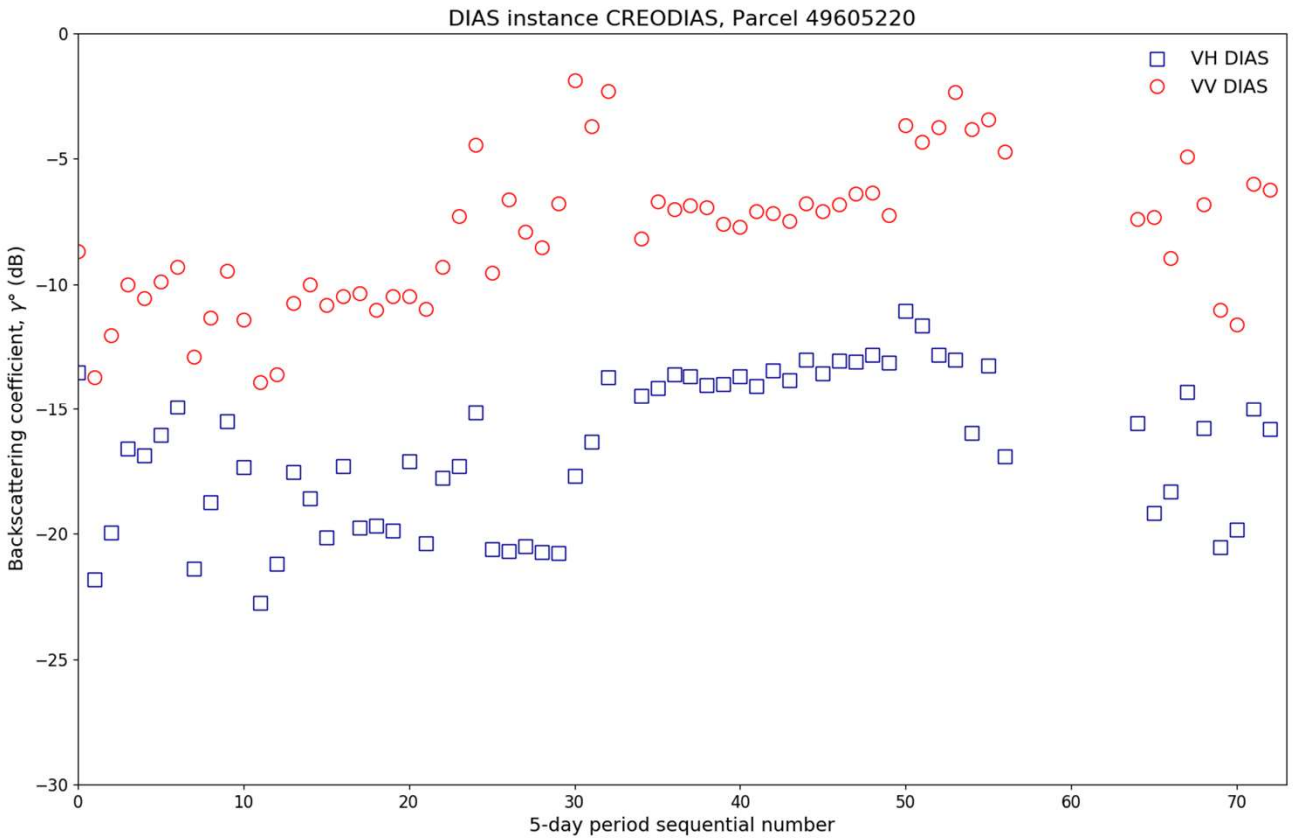
Consistency compared to Google Earth Engine

Using same s1tbx graph

Same for other DIAS

Some gaps in CREODIAS set

European Commission

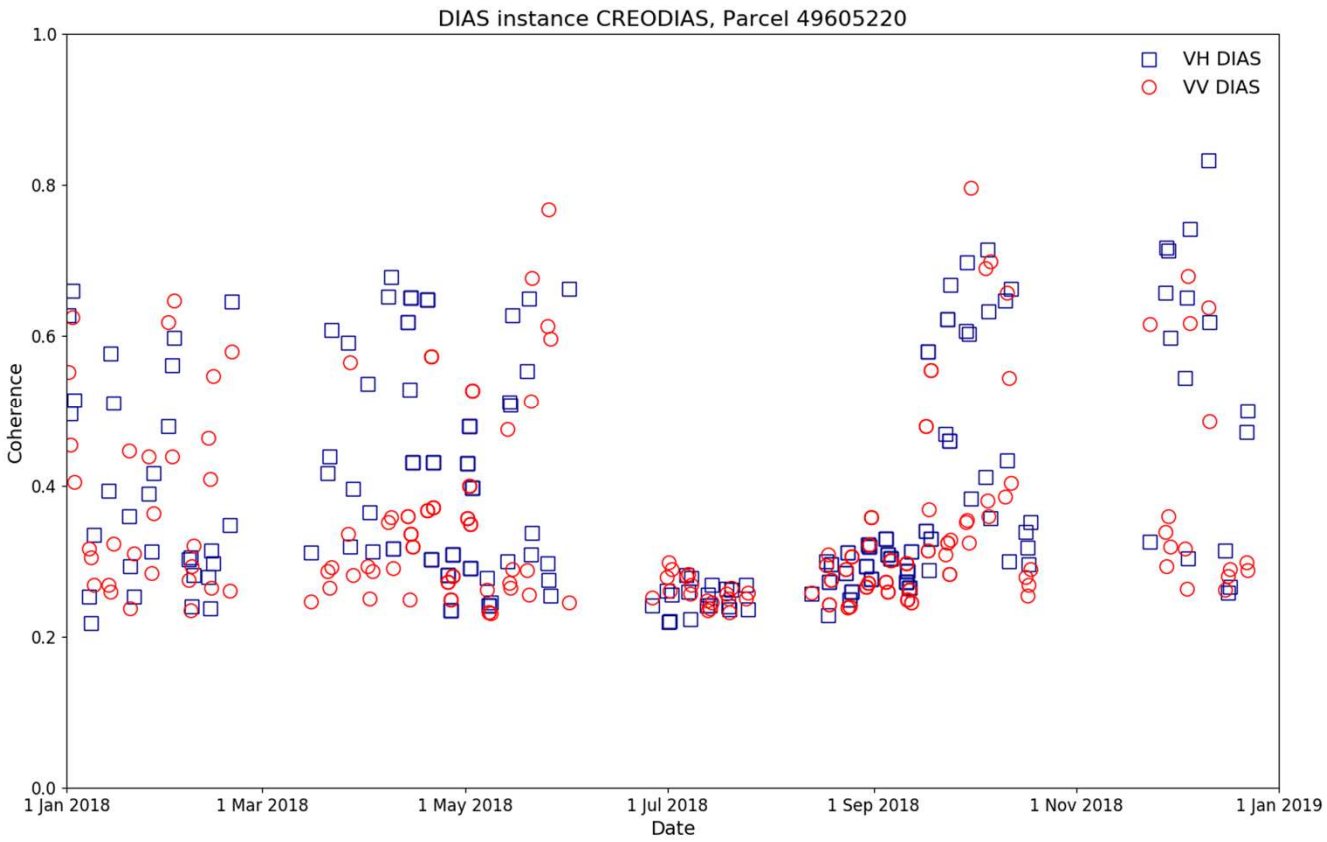


Averaged for 5 day periods

75+ million records for 2018

Input to machine learning

European Commission



CARD-COH6 per parcel

6-day coherence

250+ million records for 2018

Not available anywhere else!

Some gaps due to failed S1 orbits

European Commission

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.

European Commission

Next steps

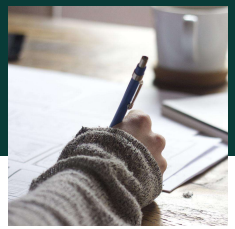
Further escalating technical solutions (kubernetes, machine learning)

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

European Commission