DIAS onboarding

IACS Autumn 2020 Webinar, 5 November 2020

JRC D5 – GTCAP Team

Scope & Status

Context

- Checks by Monitoring introduces **continuous use of Sentinel data streams** for 100% of the Member State territory.
- Copernicus DIAS advantages:
  - Access to a **consistent, complete** Sentinel data archive (push, not pull)
  - Provision of on-demand standard CARD processing
  - Access to compute resources that can (temporarily) scale to needs
  - Based on **open industry standards**, core open source modules
- Facilitates the needs for required automated processing.
- Potential for shared methodology

Experience gained with DIAS in 2019

- DIAS instances are functionally equivalent, differ in technical implementation
- Code portability is a minor issue, with 100% open source code, abstracted configurations and a core set of common modules
- Performance highly configurable on demand, suited to meet CbM demands
- But **DIAS configuration and programming requires specific expertise**
- **Need to support data access and analytics for non-expert users** (e.g. PA operators, field inspection, other practitioners)
- Closely linked to (perceived) needs of CbM practice
- 2020 DIAS activities can dive directly into operational aspects

The 2020 DIAS arrangement

- DG AGRI and DG DEFIS have arranged for new DIAS support in 2020.
- Includes the 4 ESA managed DIAS. Extended to the use of WEkEO DIAS.
- For 3 groups of CbM users with variable experience and commitments.
  a. PAs committed to CbM in 2019 (continuation)
  b. PAs to apply CbM in certain schemes in 2020
  c. Other PAs interested in pilot testing (up to 50 kkm^2)
- **"Operational"** use of DIAS IaaS, metered and reported monthly
- Actual uptake (end of Oct): way below expectations!

European Commission

Technology overview

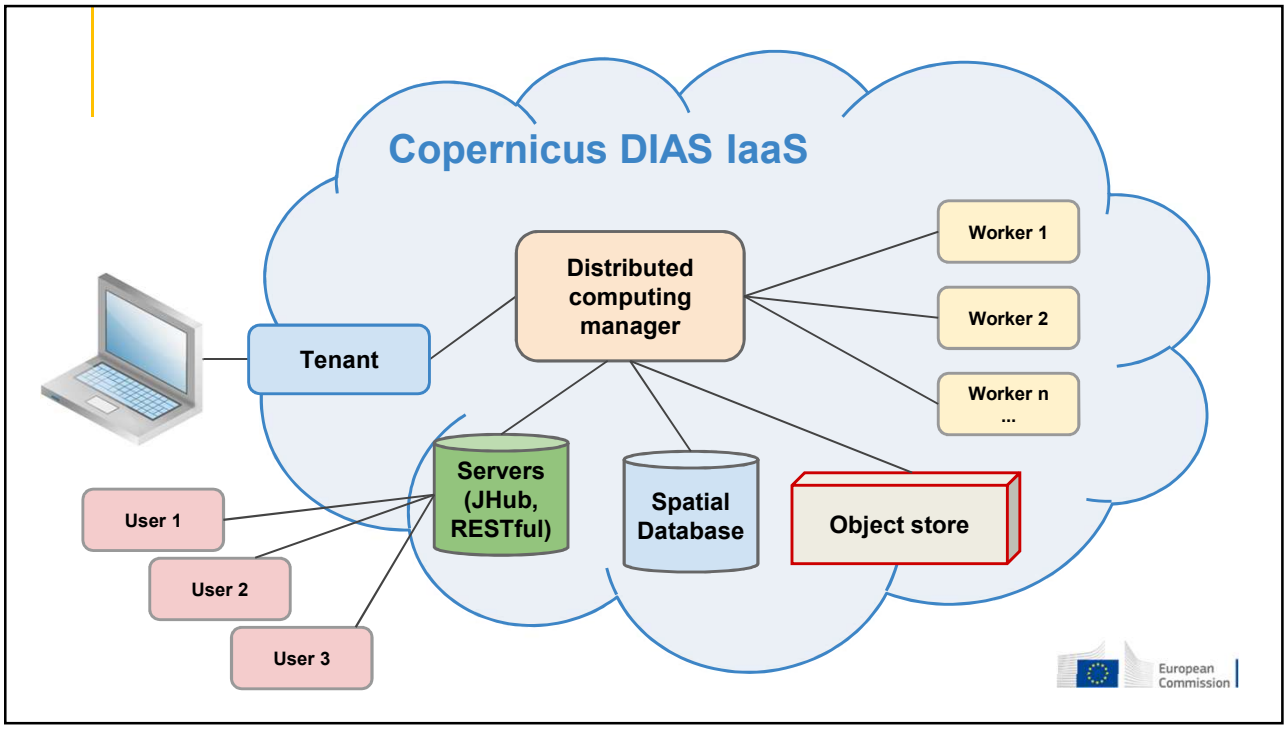
European Commission

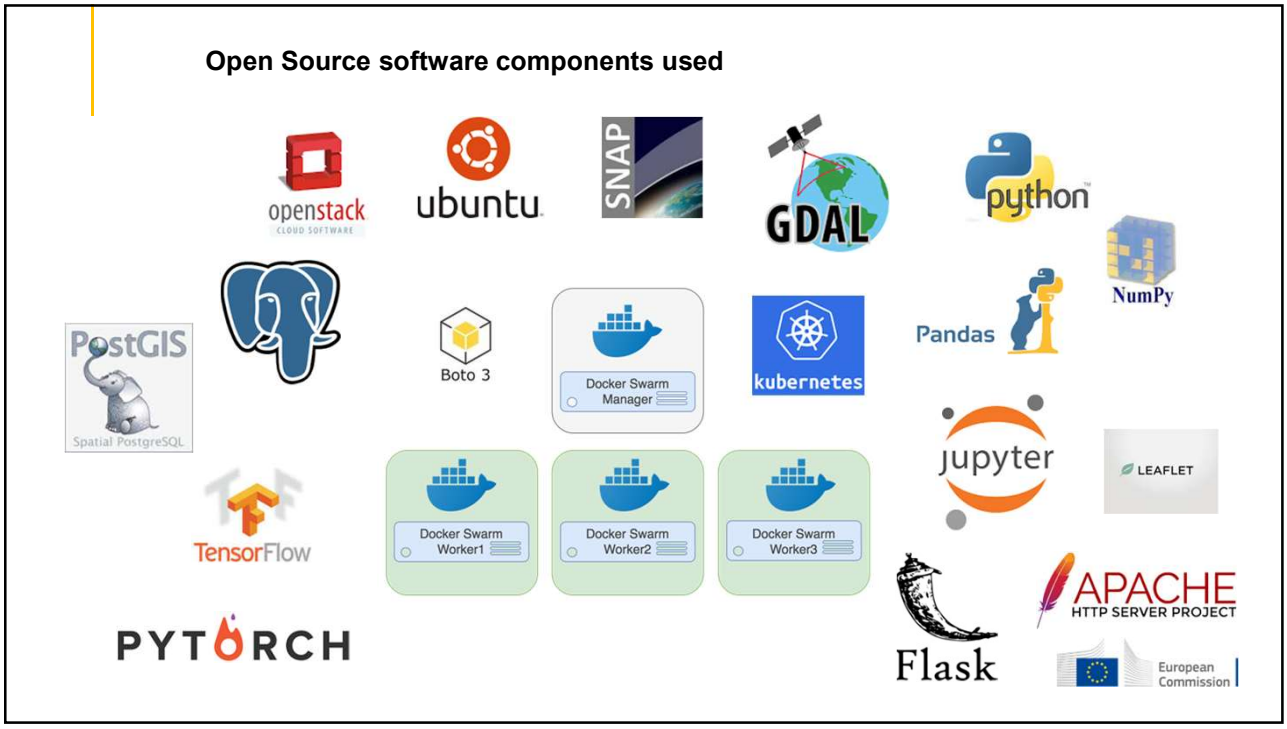
DIAS concepts (JRC view)

- Large volume processing must be fully automatic using a single standard
- Full territory application of CbM implies automated processing at parcel level
- Implies **compliant FOI cardinality**
- **Reduction is essential**, but full traceability to the source data required
- CbM markers are much more generic than “crop type” detection, can be adapted to local context and may evolve over time
- Access to data for analytics needs to serve the **typical use patterns**
- Analytics **portrayal** is up to the user, we support with adaptable code

DIAS modules (JRC view)

- Full territory Sentinel-1 CARD generation (BS, COH-6)
- Automated time series extraction of S1-CARD and S2-Level 2A
- Both in automated batch mode and interactive, on demand
- A spatial database server to control and store extraction outputs
- **RESTful** and **Notebook** server/client API to provide access and analytics
- Flexibility in deployment choices, everything is based on **Open Source**
- Modules are deployable across DIAS instances, non-DIAS cloud servers and your existing internal processing environment.


Copernicus DIAS IaaS
Distributed computing manager
Worker 1
Worker 2
Worker n
...
Tenant
Servers (JHub, RESTful)
Spatial Database
Object store
User 1
User 2
User 3
European Commission


Open Source software components used
openstack
ubuntu
SNAP
GDAL
python
NumPy
PostGIS
Spatial PostgreSQL
Boto 3
Docker Swarm Manager
kubernetes
Pandas
jupyter
LEAFLET
TensorFlow
Docker Swarm Worker1
Docker Swarm Worker2
Docker Swarm Worker3
PYTORCH
Flask
APACHE HTTP SERVER PROJECT
European Commission

DIAS figures

- Some performance figures
- PA area: 50,000 km^2, 700,000 parcels
- Annual Sentinel volume: 1500 S1 CARD-BS, CARD-C6, 4000 S2 (**~ 25 TB**)
- **Latencies**: S1 typically few hours, S2 up to 24 hours
- Extraction annual volume on 4 parallel VMs
- ~ 120 hours (2x2 bands S1, 3 bands S2) (i.e. 20 minutes/VM/day)
- Leads to ~ 60 GB time series database tables

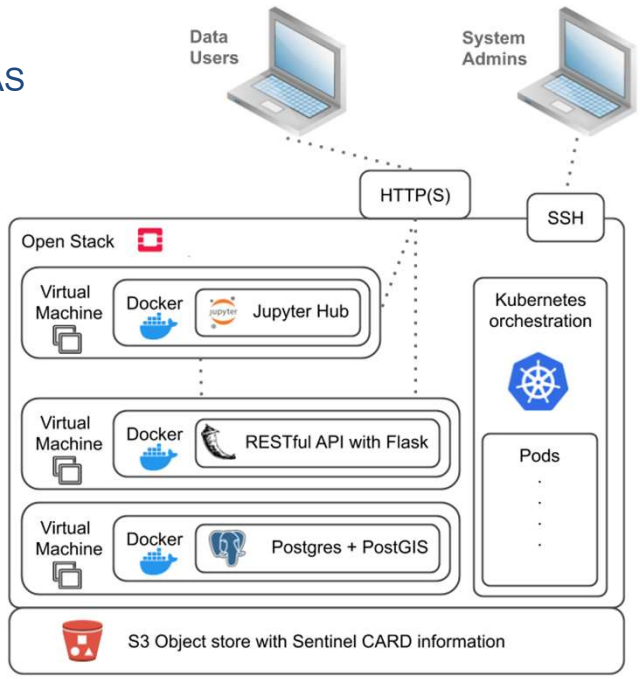
Interfaces

DIAS roles and use patterns

- **ICT expert** maintains IaaS and the required server components
- **Data analytics expert** programs and runs reduction functions and core analytics (e.g. extraction, machine learning)
- **Data users** get access via server API (e.g. RESTful, JHub, map server)
- Typically for “yellow cases”, i.e. to inspect identified outliers
- **RESTful** supports time series analysis (fast), sub-image selection (slow)
- Can be “consumed” in scripts, automated reports, Jupyter notebooks, on your mobile phone, etc.

European Commission

Simplified diagram of DIAS components

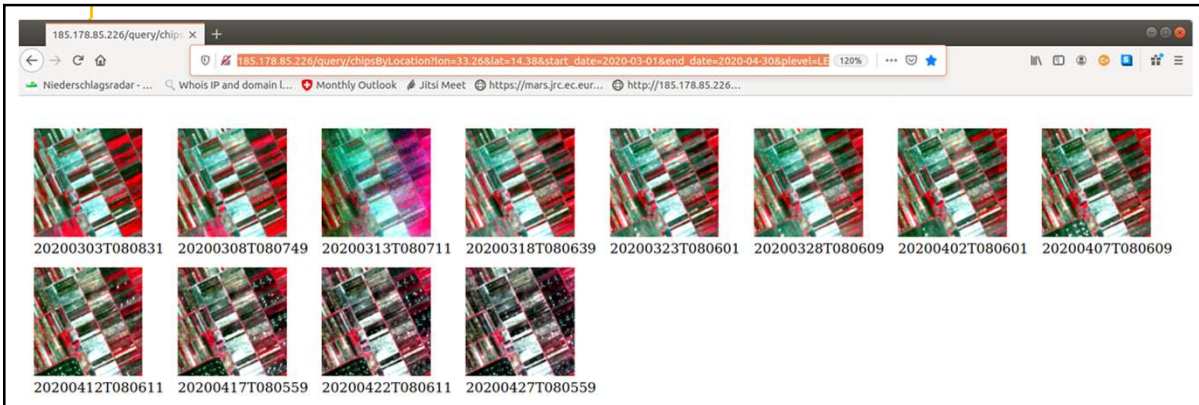


European Commission

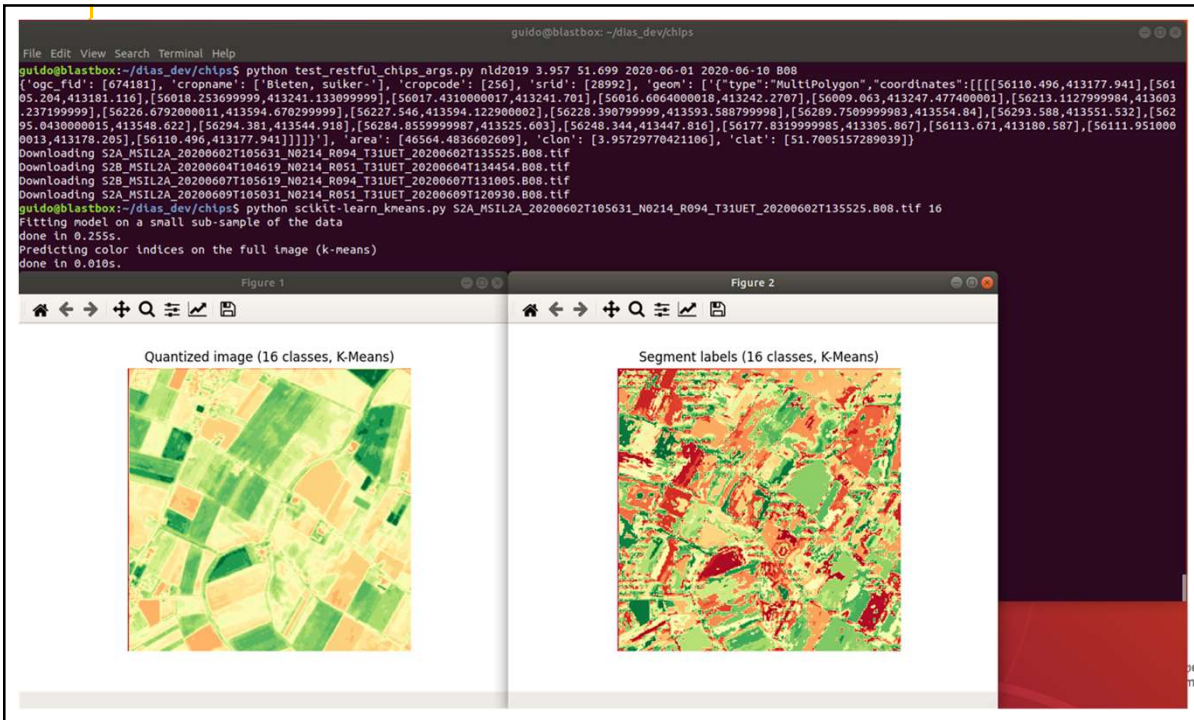
DIAS RESTful services

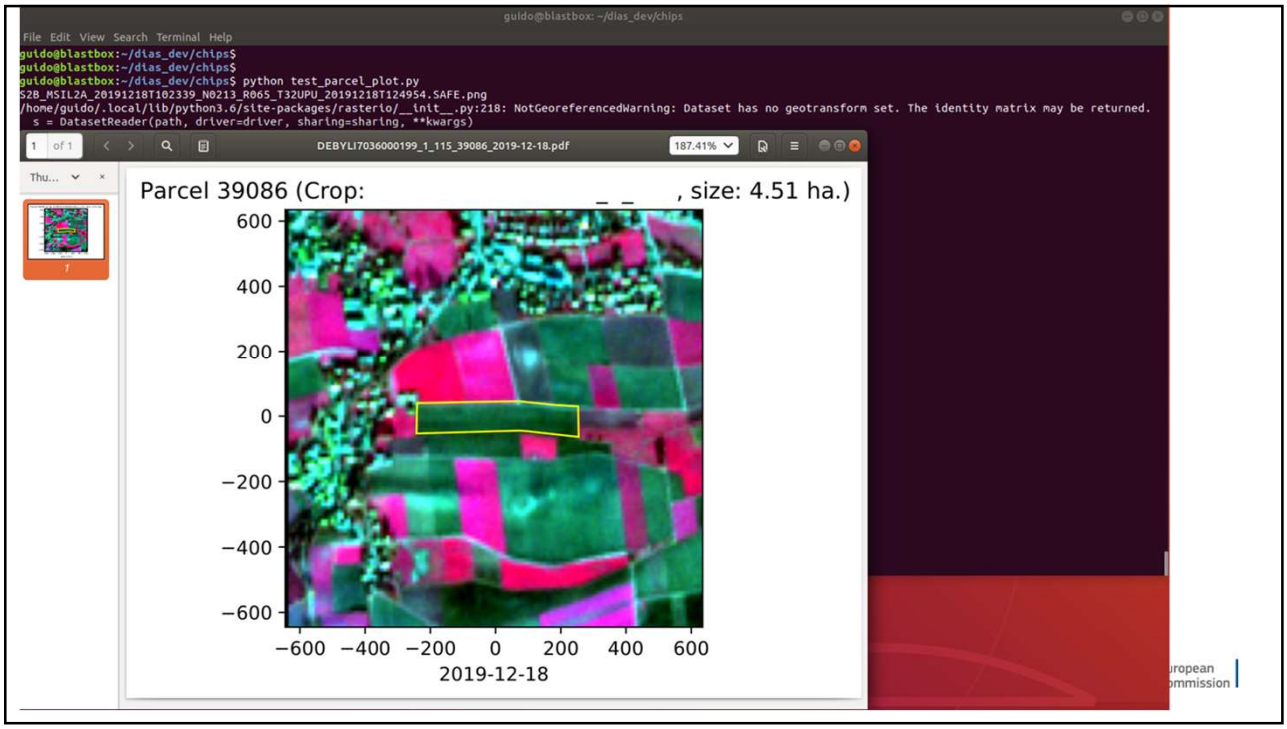
- Run in a Flash server (python centric)
- Basic information retrieval (**parcelByLocation**)
- Fast parcel time series statistics (**parcelTimeSeries, parcelPeers**)
- (Slower) image chip selection, for visualization (**chipByLocation, backgroundByLocation**)
- idem, but full resolution GeoTIFFs (**rawChipByLocation**, **rawChipsBatch**).
- All come with example scripts on how to “consume” client-side (see WIKI)
- Later presentations today demonstrate practical use cases

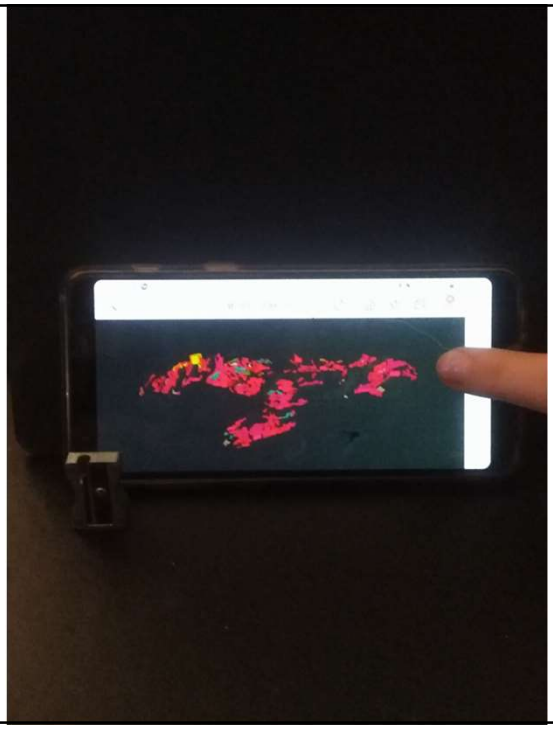
RESTful demos



Sentinel-2 chip visualization in browser, using **chipsByLocation.** Location in Sudan, to demonstrate global scope.




guido@blastbox: ~/dias_dev/chips
File Edit View Search Terminal Help
guido@blastbox:~/dias_dev/chips$
guido@blastbox:~/dias_dev/chips$
guido@blastbox:~/dias_dev/chips$ python test_parcel_plot.py
S2B_MSIL2A_20191218T102339_N0213_R065_T32UPU_20191218T124954.SAFE.png
/home/guido/.local/lib/python3.6/site-packages/rasterio/__init__.py:218: NotGeoreferencedWarning: Dataset has no geotransform set. The identity matrix may be returned.
s = DatasetReader(path, driver=driver, sharing=sharing, **kwargs)
DEBYLI7036000199_1_115_39086_2019-12-18.pdf
Parcel 39086 (Crop: , size: 4.51 ha.)
600
400
200
0
−200
−400
−600
−600 −400 −200 0 200 400 600
2019-12-18



Jupyter Notebook demos

Other relevant activities

Third party CbM developments of interest

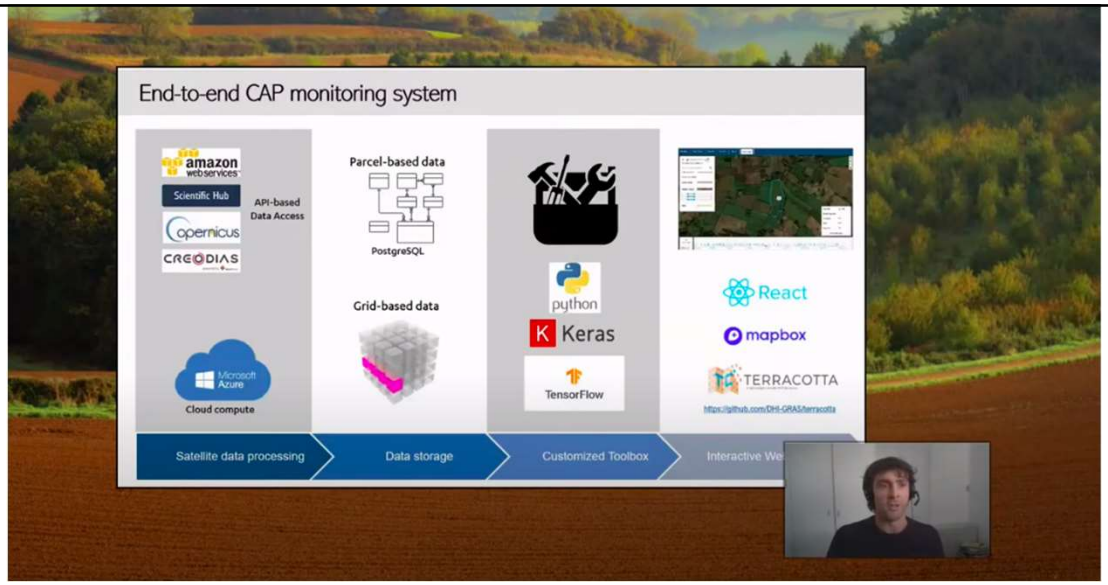
- Appearance of cloud-based CbM solutions in the market
- Adopting the modular approach, or migrating to it (e.g. SEN4CAP)
- Or providing relevant analytical modules implemented in open source

- PA preferring a local cloud + HPC solution (e.g. Castilla y León)
- JRC participates in European Open Science Cloud “Early Adaptor Project”

- Next slides are a few anecdotal evidences
- Does not pretend to be complete, and is not in any way preferential
- Check out details at leisure with the links provided

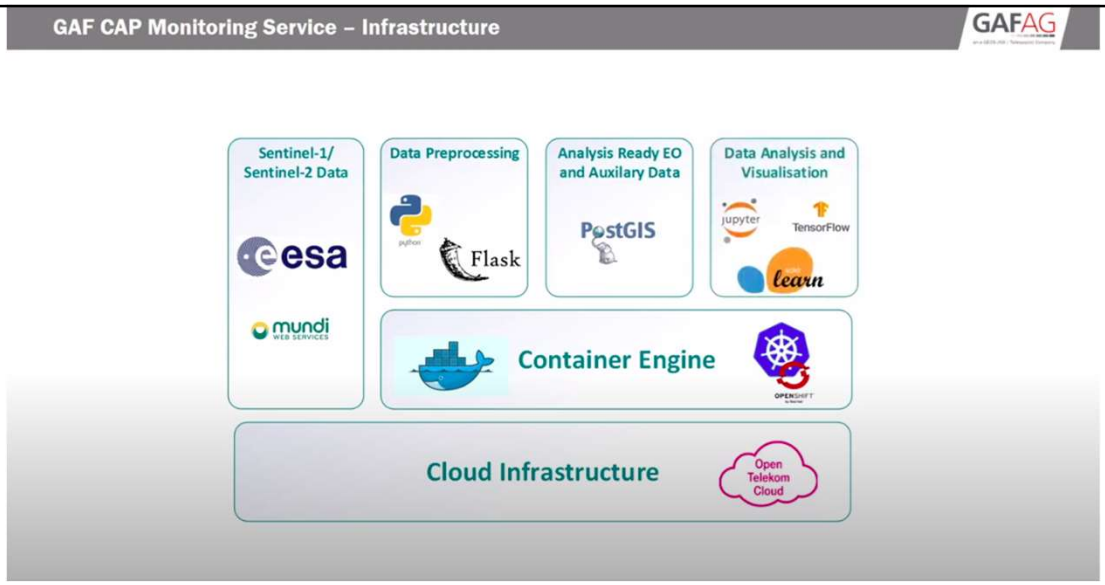
- Happy to receive additional references

European Commission



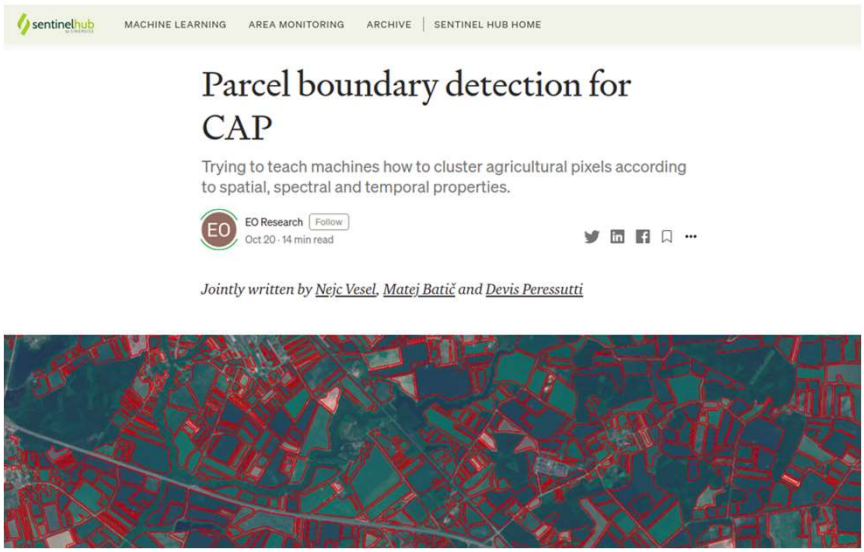
- DHI Gras presentation at ESA “EO4Agriculture under pressure” webinar
- https://www.youtube.com/watch?v=SYPRE_Lw6e4&start=226&end=1200

European Commission



- GAF presentation at ESA “Φ Week 2020” webinar
- https://www.youtube.com/watch?v=8pnwjXPsxd8&start=75&end=174 (hit ‘m’ for no audio)

European Commission



- Sinergise open sourcing “parcel boundary detection” (as part of NIVA)
- https://medium.com/sentinel-hub/parcel-boundary-detection-for-cap-2a316a77d2f6

European Commission

Outlook

Next steps

- DIAS 2020 onboarding arrangement will extend into 2021
- More effort needed to tackle entry hurdles
- Substantial role for technical services in Member States

- Discussion on S1 CARD processing with DEFIS (at least EU wide)
- DEFIS review of Copernicus DIAS contracts (2021)
- Digital Europe programme for EOSC, EuroHPC, European Data Spaces

- JRC support to install & configure, but requires active technical support PA
- JRC to develop further documentation, code, webinar-style training
- JRC open to discuss optimization steps with DIAS providers (data formats, scalability, alternative processing, data integration)

Take home messages

- Cloud services are a compute utility, DIAS is specialized for Sentinel data use
- The JRC modular approach provides flexibility in deployment choices
- Technical expertise required to “cloudify” the CbM specific data use

- Copernicus infrastructure evolution important (CARD, DIAS future, DestinE)
- Engage with your national Copernicus Program Coordinators!

- Development can now move towards systematic marker documentation
- Integrate more sophisticated processing, link in open MS data sets
- Scaling and robustness of server solutions
- Consider other operational aspect of CbM (e.g. support to GSAA, farmer warning, traffic light management)

References

DIAS documentation and code

- JRC DIAS Wiki Pages and code repository
- Needs username and password (ask Kostas)

  http://jrc-ntb.vm.cesnet.cz/DIAS/jrc-dias/wiki/0.-DIAS-for-CAP-Checks-by-Monitoring

- Includes a full description of RESTful services + code examples (but runs against EOSC pilot data sets in NL some ES)
- We maintain some other resources, e.g.

  https://hub.docker.com/r/glemoine62/dias_py

European Commission

Q&A

guido.lemoine@ec.europa.eu
konstantinos.anastasakis@ext.ec.europa.eu

European Commission