The European Commission’s science and knowledge service

Joint Research Centre

European Commission

User Communities: Earth Observation

EGI Workshop on Data Management, Amsterdam, 3-4 July 2019

Guido Lemoine (guido.lemoine@ec.europa.eu)

European Commission

Rationale

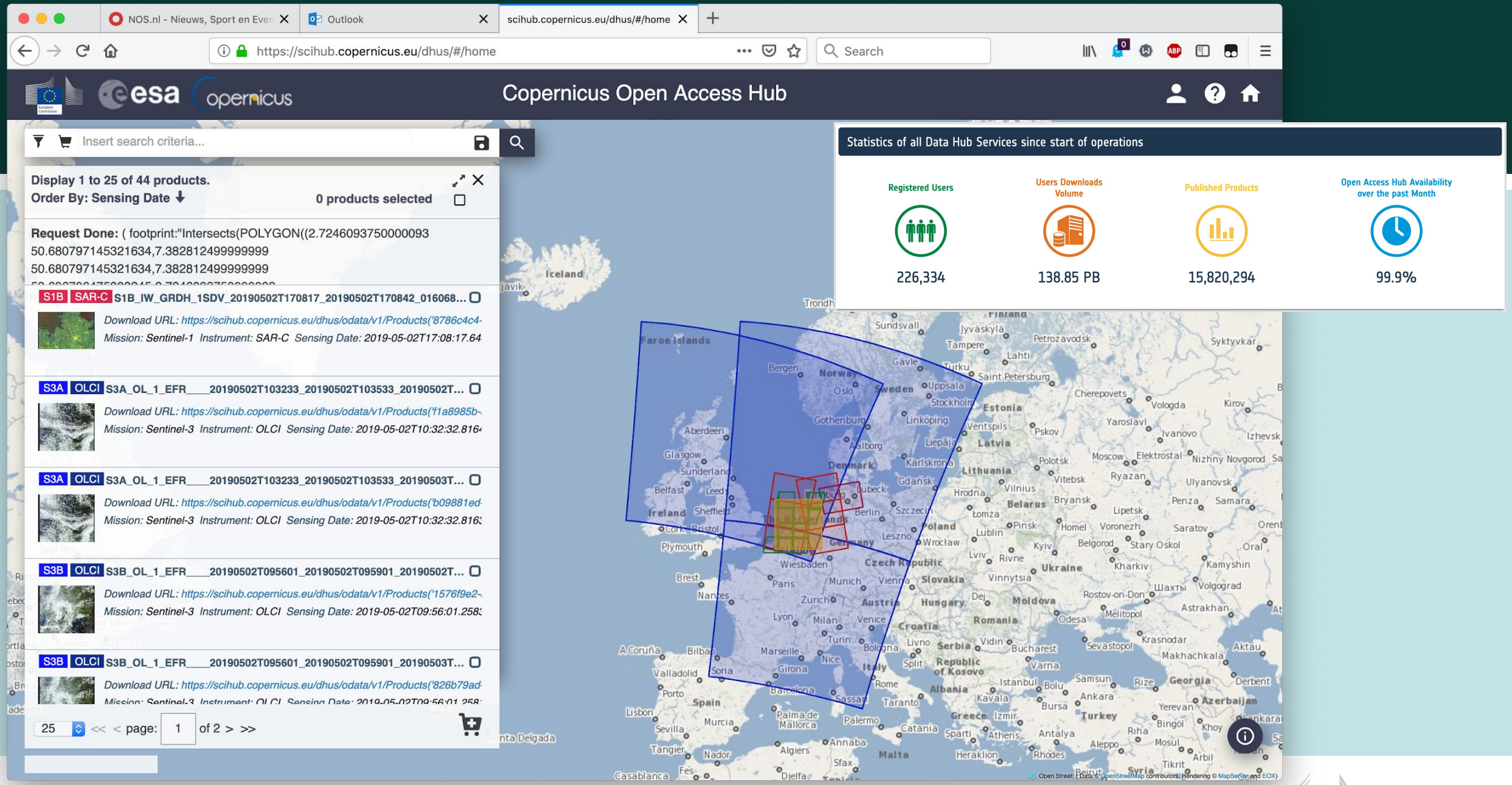
The European Union's Copernicus program is the global leader in Earth Observation for environmental monitoring.

Data from 4 different sensors types (7 operational platforms) are provided under a full, free and open data license. Currently data rate is ~ 6 PB/year.

Includes Sentinel-1 and -2 high resolution (10 m) global coverage, with 6- and 5-day revisit. Lower resolution S-3 and S-5P with near-daily coverage.

Data accessed by 1000s of users via ESA ScienceHub, downloaded 10-fold.

Leading to dramatic scale up in data uptake, spatially, temporally and spectrally.



European Commission

Rationale (2)

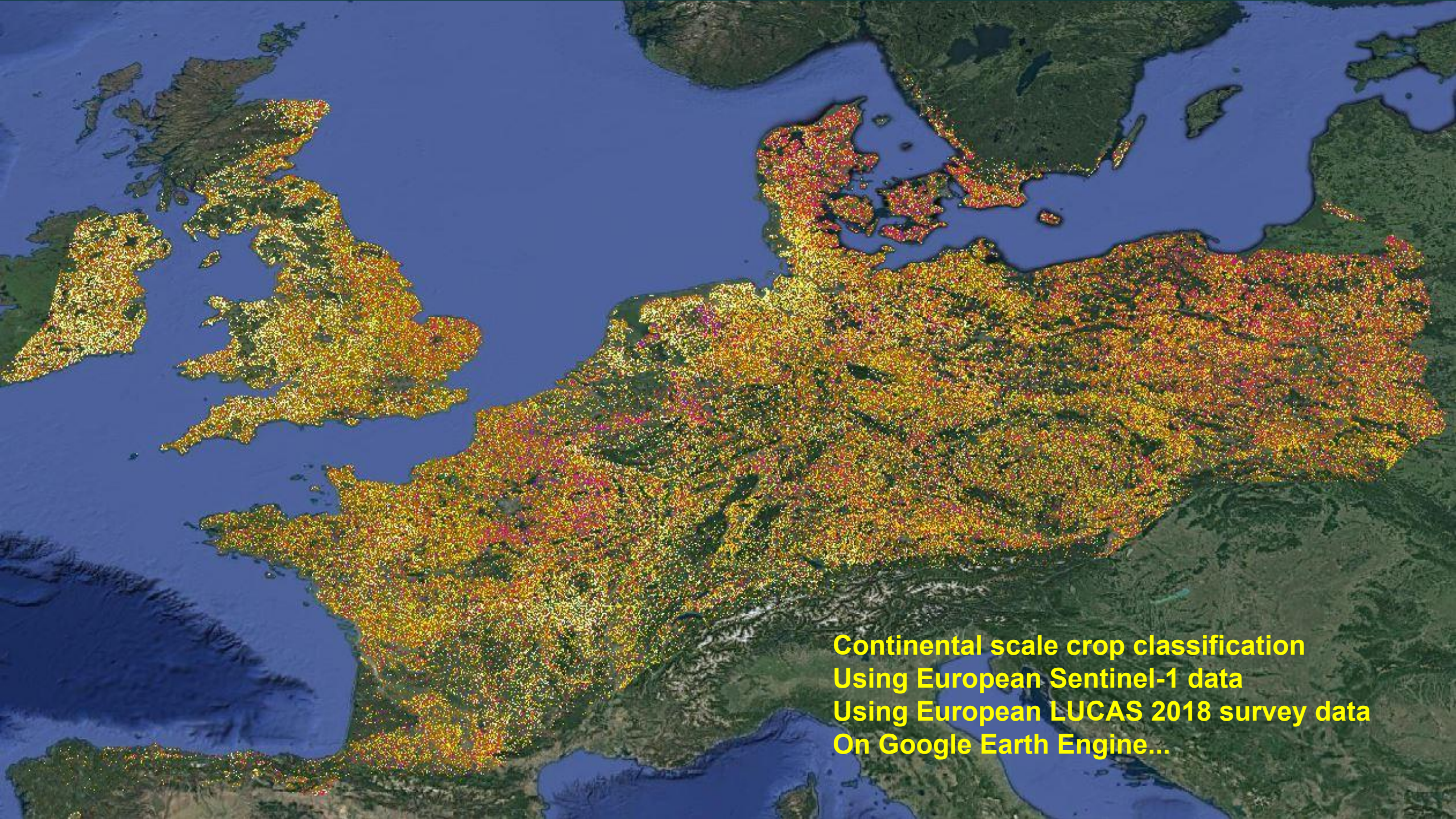
An obvious “Big Data Analytics” domain, but **poorly served by European capacity**

Currently, BDA (much) better served by AWS and **Google Earth Engine**.

European solutions tends to be fragmented, across thematic use, national audiences, user sectors. No concerted science community support.

European solutions are not state-of-the-art, don’t scale well, are not user friendly.

Europe does not have a consolidated plan to **maintain the full data archive**!


Continental scale crop classification
Using European Sentinel-1 data
Using European LUCAS 2018 survey data
On Google Earth Engine...

Background

The “Earth Observation” [science] community is an applied research community.

Paradigm change (frequent open data) has created huge capacity bottleneck, leading to rapid uptake of novel approaches in Big EO Data Analytics.

Primarily about fast upscaling, in time, space, across sensors of large code base of existing algorithms. Deeper integration of existing and new open reference data.

Our role is to stimulate uptake of Copernicus data in scientific and public use, linked to EU policy goals.

We have [use] experience with most major initiatives, initial exposure to EOSC.

No XDC service use yet, but XDC template fits many requirements.

Users

The “Earth Observation” [science] community spans a wide range of disciplines (e.g. geophysics, agriculture, land use, marine, atmosphere, economic sciences)

Typically with a short path to data applications (public and private users)

1000s of European users (GEE has 50K worldwide users)

Bulk of users require “application ready data” used in batch and interactive analytics based on open standards and open source algorithms.

Integration into tag-on “decision support” is a typical use case (e.g. land use change, crop status, water quality, impact enumeration, etc.)

System validation against external alternatives (data consistency). TBD.

Current status

Data ingestion, discovery and download services exist

(European) data storage is fragmented, usually only partial archive (CloudFerro/CREODIAS by far the largest). ESA will not maintain LTA.

Many open source algorithm libraries, but typically on per image basis. Large variety, but some common basis (e.g. GDAL, PostgreSQL/Postgis, python). Best-of-class scalable solution (GEE) is not open (server side)

e-Infrastructure should provide optimal data indexing, transparent resource marshalling, smart caching, massive parallel (micro-) tasking, vector/raster integration, advanced visualization support, open standards based, open source (both server and client side)

JRCDIAS - Google Drive | EGI_EOSC_4_EOSC - Google Si | Exercise4 - Earth Engine Code

https://code.earthengine.google.com

Google Earth Engine

Search places and datasets...

Help | gglemoine62

Scripts | Docs | Assets

ee.Classifier

- ee.Classifier.cart(*crossvalidationFactor, maxDe...*
- ee.Classifier.continuousNaiveBayes(*lambda*)
- ee.Classifier.decisionTree(treeString)
- ee.Classifier.gmoLinearRegression(*weight1, we...*
- ee.Classifier.gmoMaxEnt(*weight1, weight2, epsi...*
- ee.Classifier.ikpamir(*numBins, learningRate, ep...*
- ee.Classifier.minimumDistance(*metric*)
- ee.Classifier.naiveBayes(*lambda*)
- ee.Classifier.pegasosGaussian(*rbfGamma, loss...*

Exercise4 | Get Link | Save | Run | Reset

```
Imports (1 entry)
  var geometry: Point (5.55, 52.54)
var brp2017 = ee.FeatureCollection('users/gglemoine62/BRP_gewasperceler

brp2017 = brp2017.map(function(f) { return f.set({'id': f.id(), 'area':

brp2017 = brp2017.filterMetadata('area', 'less_than', 10000000) // Remo

var gemeenten = ee.FeatureCollection('ft:1B3v8wxCkO1aGd8jF4byitKEjolHv(

//var aoi = gemeenten.filter(ee.Filter.inList('gemnaam', ['Ede', 'Wager
//var aoi = gemeenten.filter(ee.Filter.inList('gemnaam', ['De Ronde Ver
var aoi = gemeenten.filterMetadata('first_prov', 'equals', 'Flevoland')

brp2017 = brp2017.filterBounds(aoi)
```

Inspector | Console | Tasks

Use print(...) to write to this console.

FeatureCollection users/ggle... JSON
- type: FeatureCollection
- id: users/gglemoine62/BRP_gewas...
- version: 1499814336379000
- columns: Object (0 properties)
- features: List (3 elements)
- properties: Object (1 property)
  - system:asset_size: 388209308

Point (682621.61, 5780280.30... JSON

European Commission

Plans for the workshop

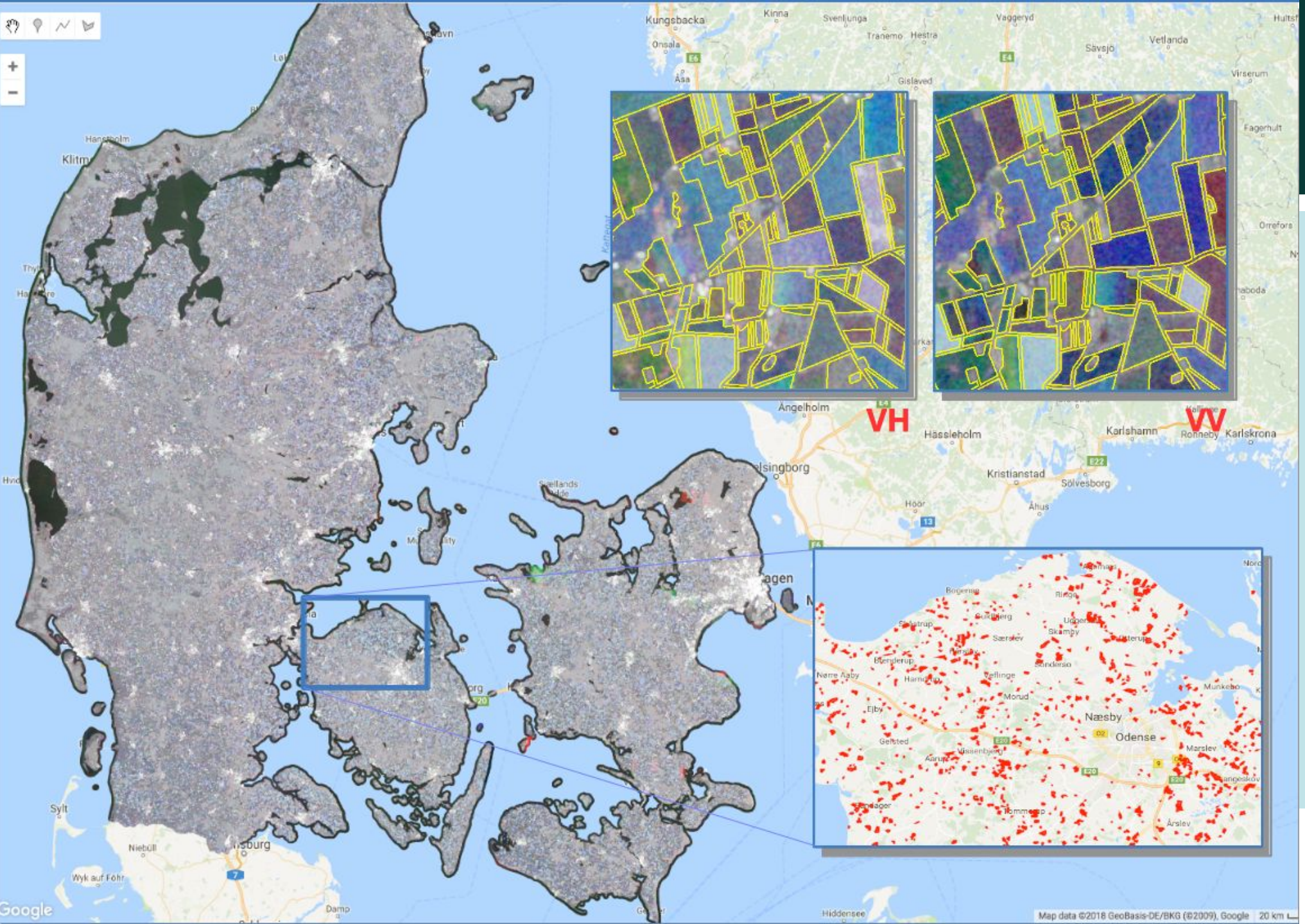
What is the relevant state-of-the-art in other structured data domains?

How realistic is it to emulate a GEE-like environment, to what extent?

Which projects, solutions, libraries are mature [partial] solutions to consider?

What kind of developer communities need to be approached, what mechanisms should be used (Horizon Europe, code sprints, sponsored, etc.)

Ultimately, can we drag EO data out of the space domain and into the “societal benefits” domain



S1 provides consistency (~200 images per year)

500K agricultural parcels with declared practice

Machine learning applied to S1 time series

Splits 95-5 “outliers”

Developed in GEE, now migrated to the Copernicus DIAS

A EuroGEOSS demonstrator (+EOSC)

European Commission

Thanks!

Any questions?

Contact guido.lemoine@ec.europa.eu

European Commission