



**Large scale mapping of human settlements from Earth
Observation data with JEO-batch of the
JRC Earth Observation Data and Processing Platform**



C. Corbane, D. Rodriguez, Sabo F., P. Politis and V. Syrris



Joint Research Centre, Disaster Risk Management Unit

**Contacts: jrc-ghsl@ec.europa.eu
Christina.corban@ec.europa.eu**

Geo-information for analysing human settlements



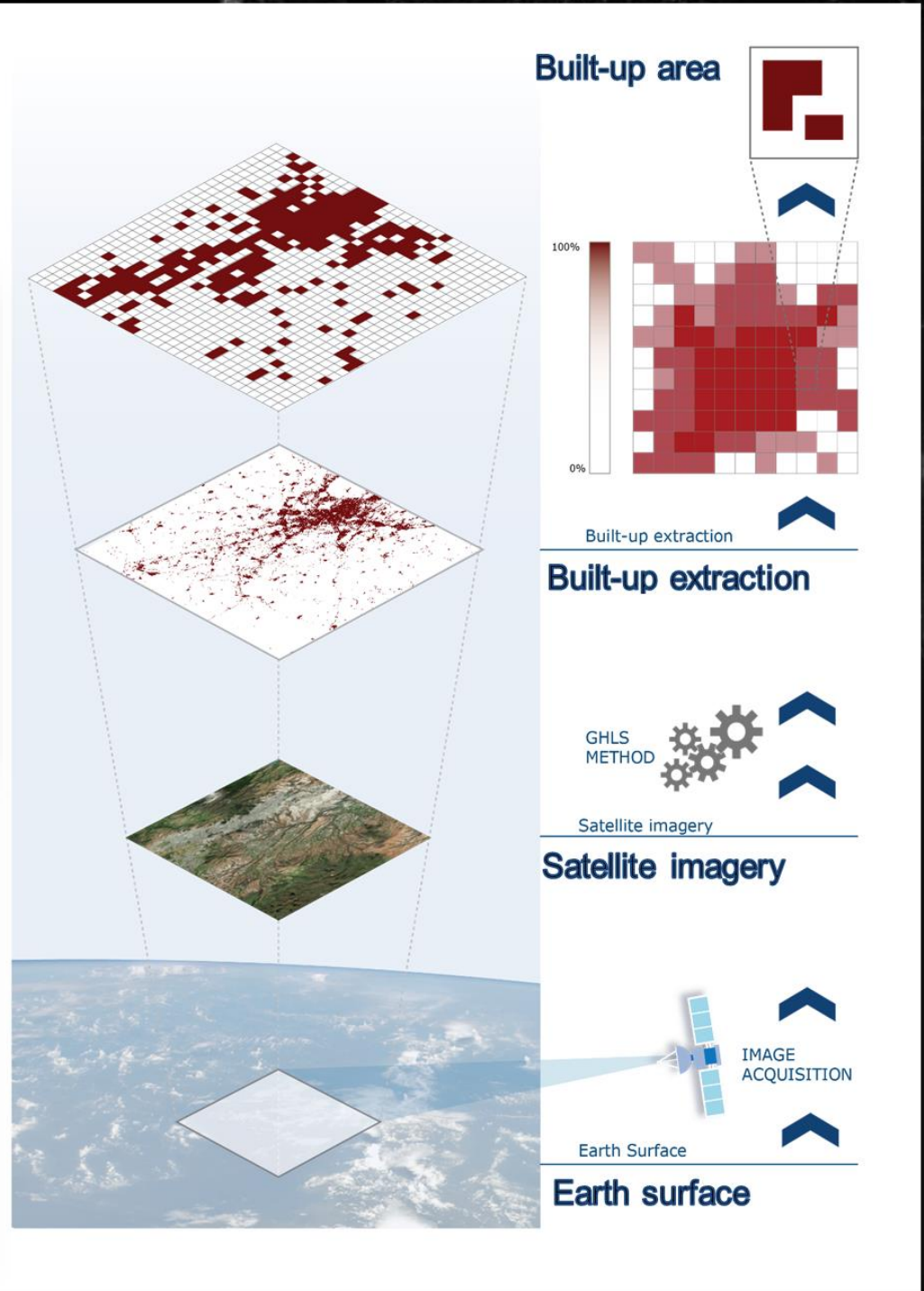
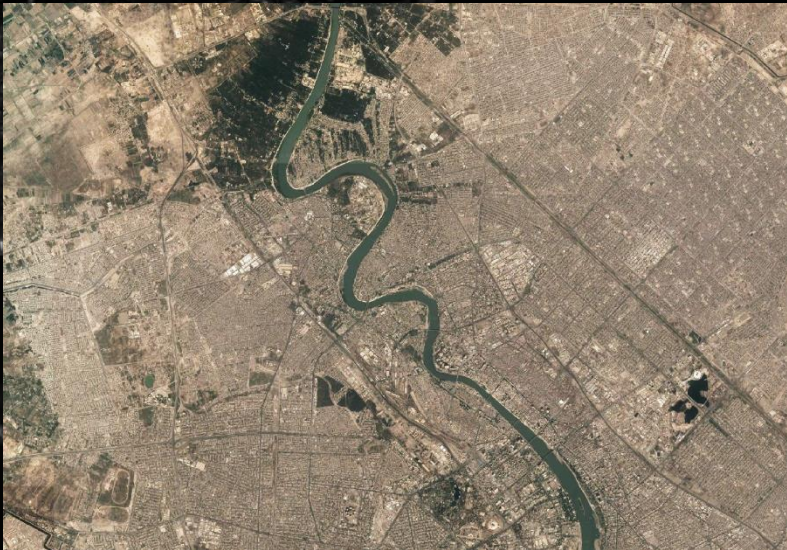
Required for modeling, analysis, policy-making (i.e. producing indicators and monitoring targets):

- Sendai Framework for DRR 2015-2030
- 2030 Agenda for Sustainable Development
- COP 21 Paris Agreement on Climate Change



- **Exposure** (to hazards, pollutants,...)
- **Access** (to resources, services,...)
- **Impacts** (from natural disasters, environmental change,...)

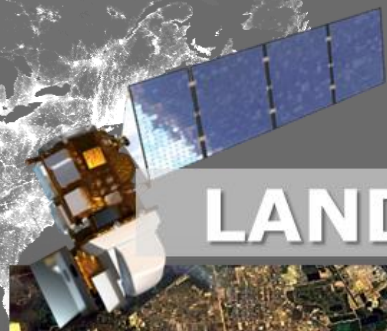
From Earth's Surface... to Pixels... to Built-up areas





What we detect: "built-up area" = all spatial units (30x30m Landsat, 10x10m Sentinel) where a roofed building or part of a building can be recognized

Global Scale Human Settlement Maps



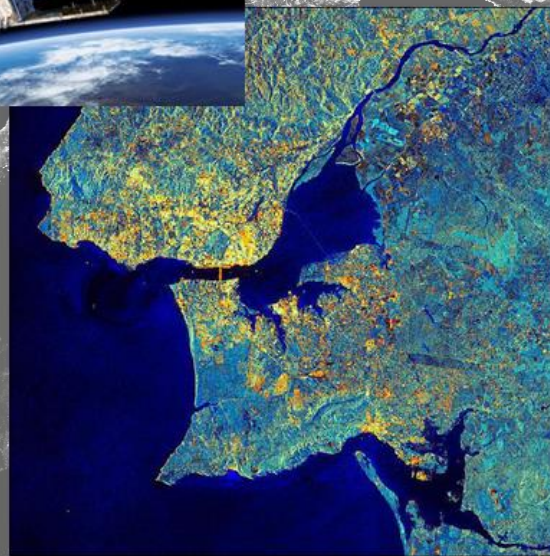
LANDSAT



**Multitemporal
(1975-1990-2000-2014)**



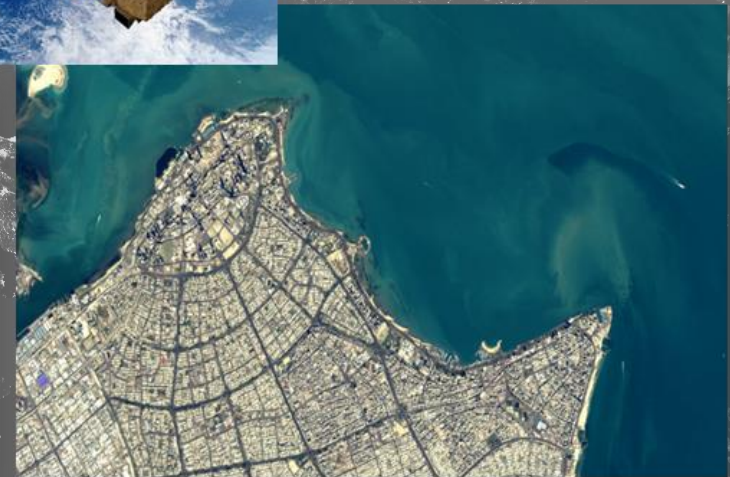
SENTINEL-1



Reference year 2016



SENTINEL-2



Reference year 2018

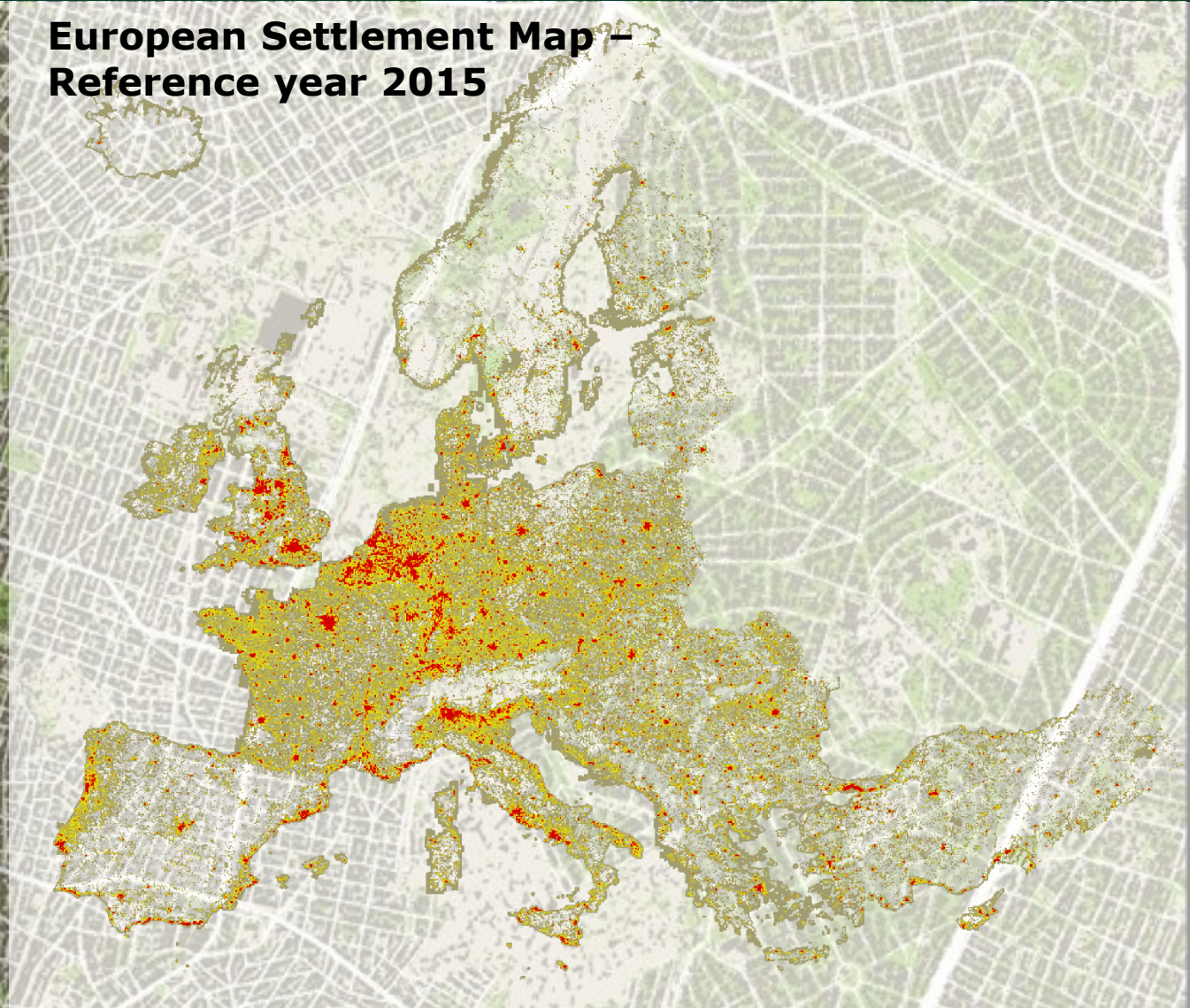


European Scale Human Settlement Maps

**Pan-European Coverage of
Very High Resolution Satellite
Imagery (2 meters)**



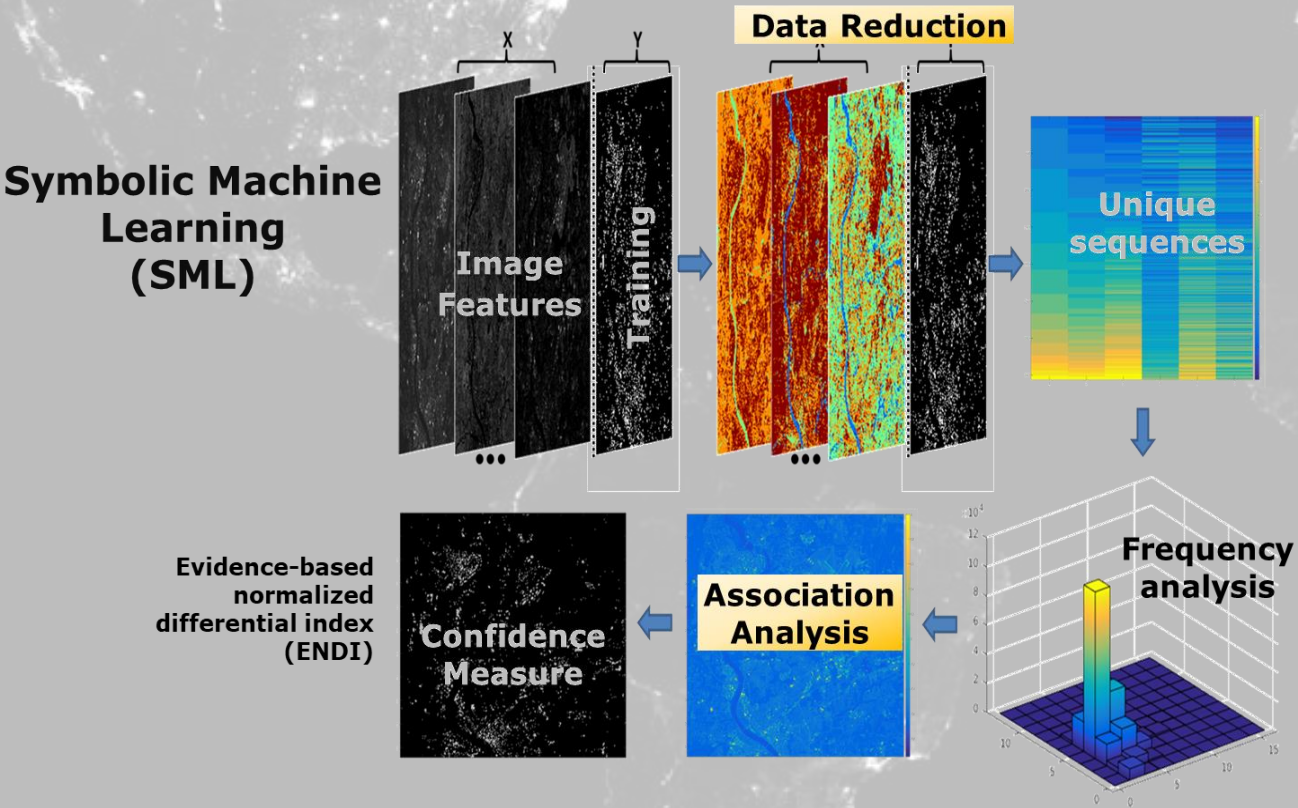
**European Settlement Map –
Reference year 2015**



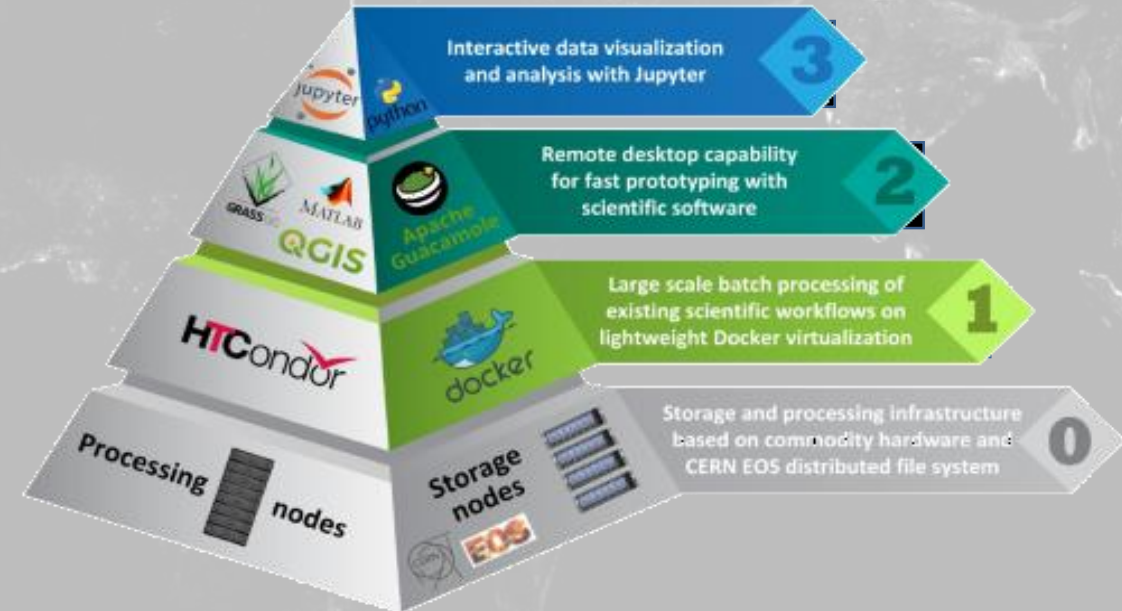
Machine Learning tools

Big Data Infrastructure

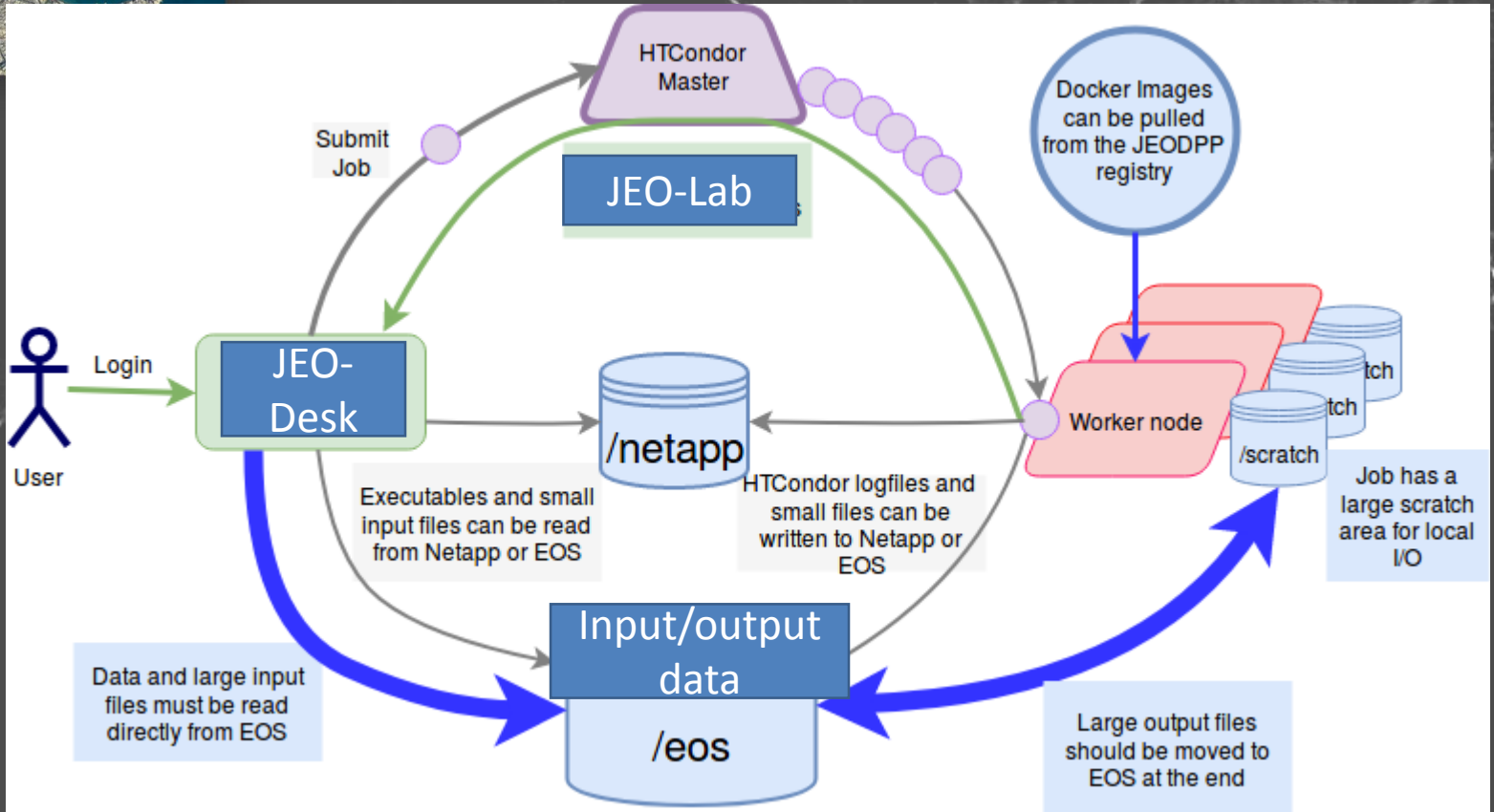
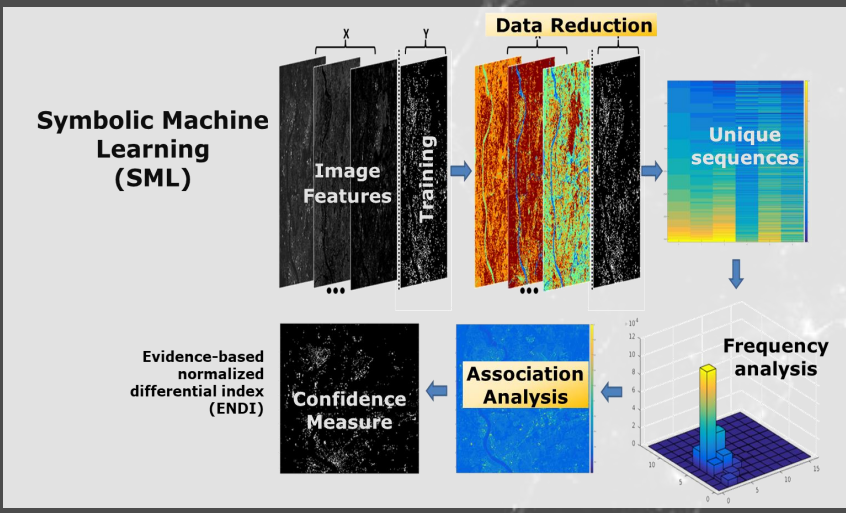
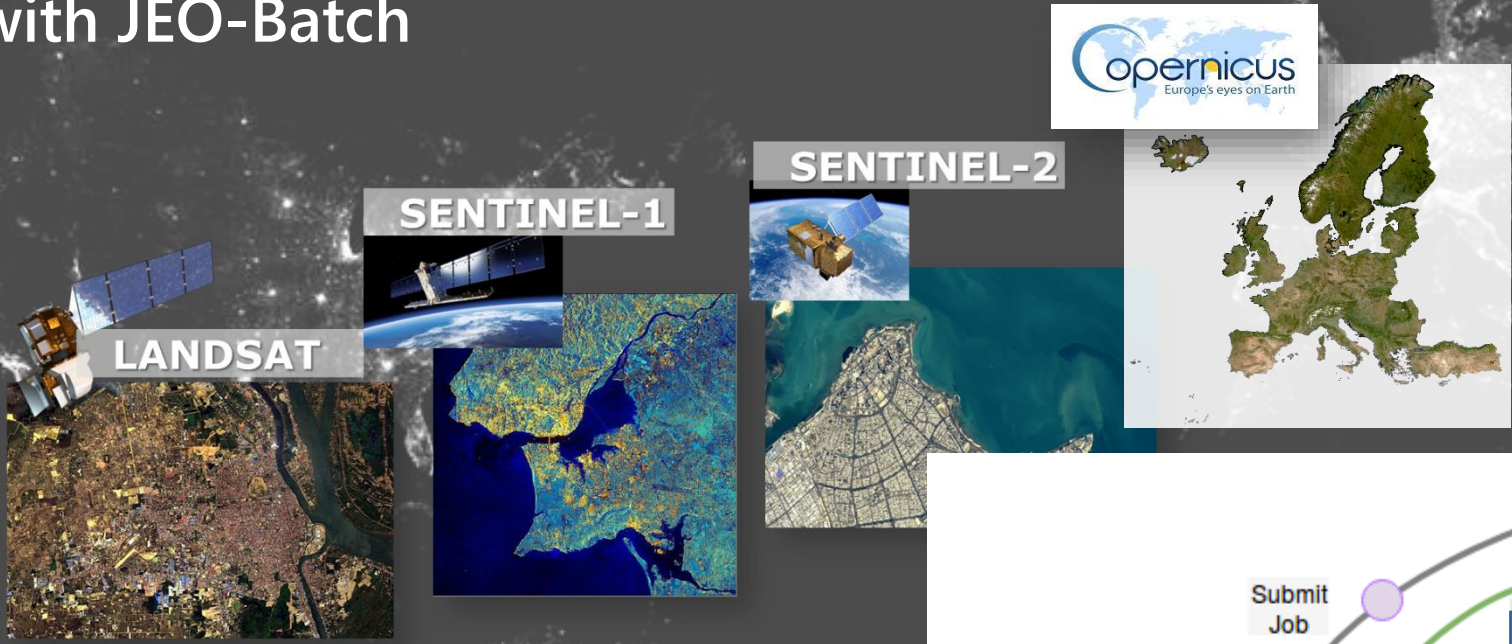
JRC Symbolic Machine Learning



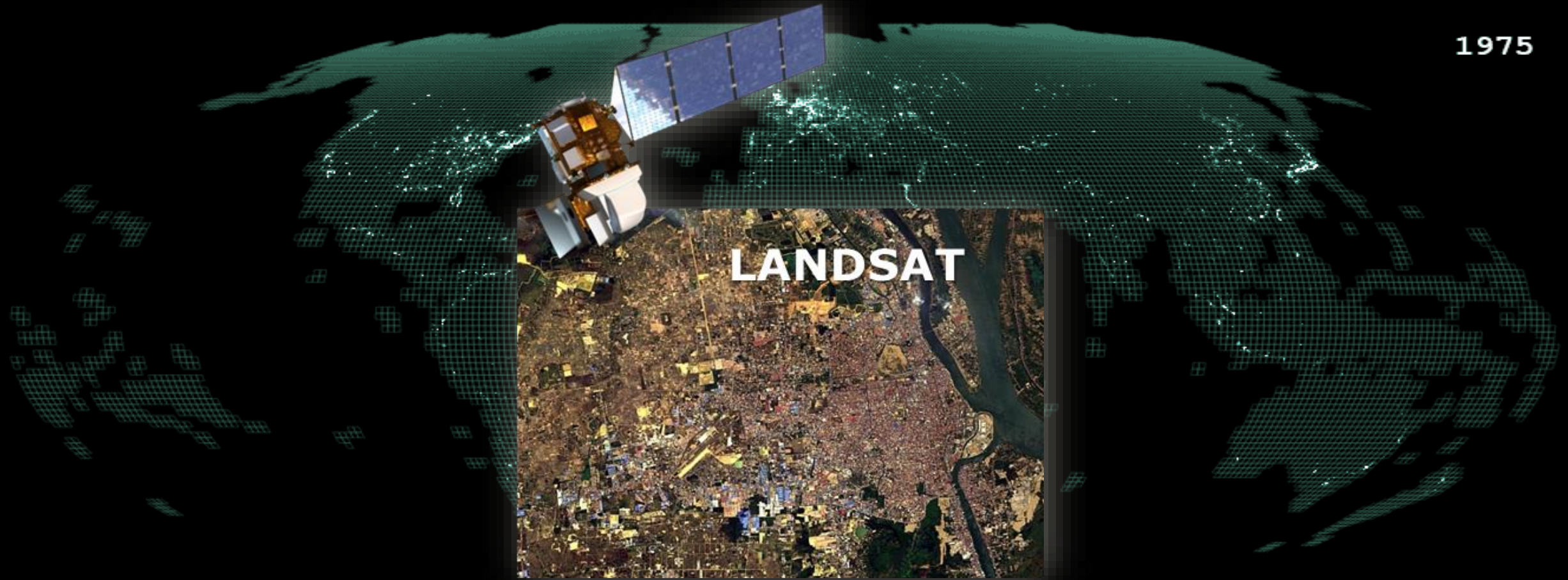
JRC Big Data Platform JEODPP



Large Scale Processing of satellite imagery: SML deployment on JEODPP with JEO-Batch



GHSL Landsat Multitemporal in 2016



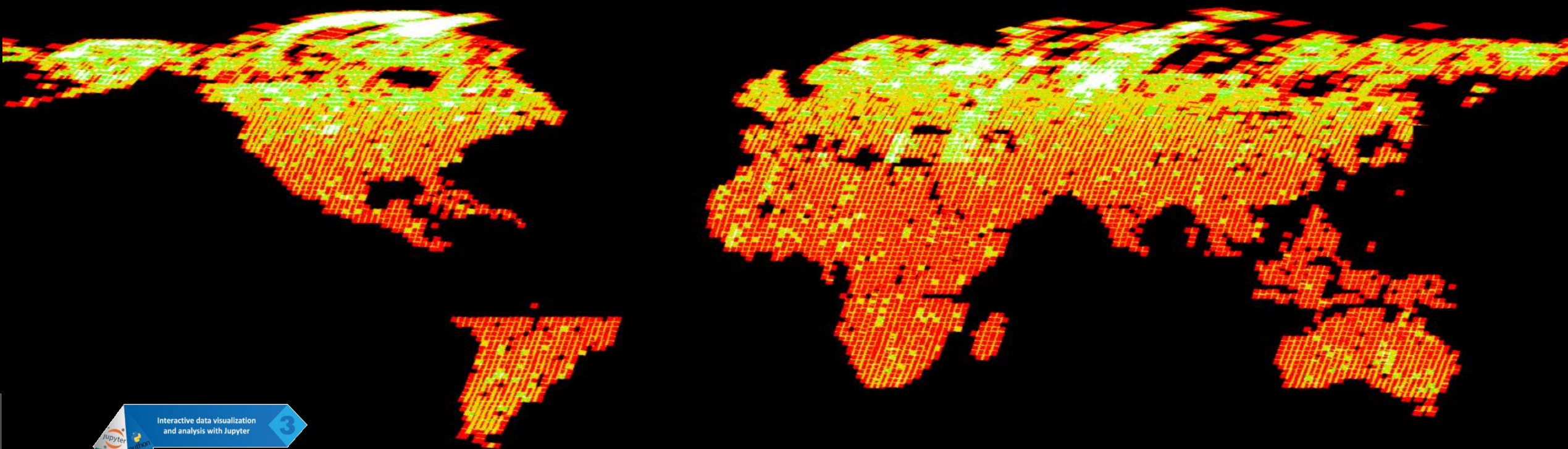
First available multitemporal assessment of built-up areas

Global coverage of Landsat data



GHS-BUILT

2019/01

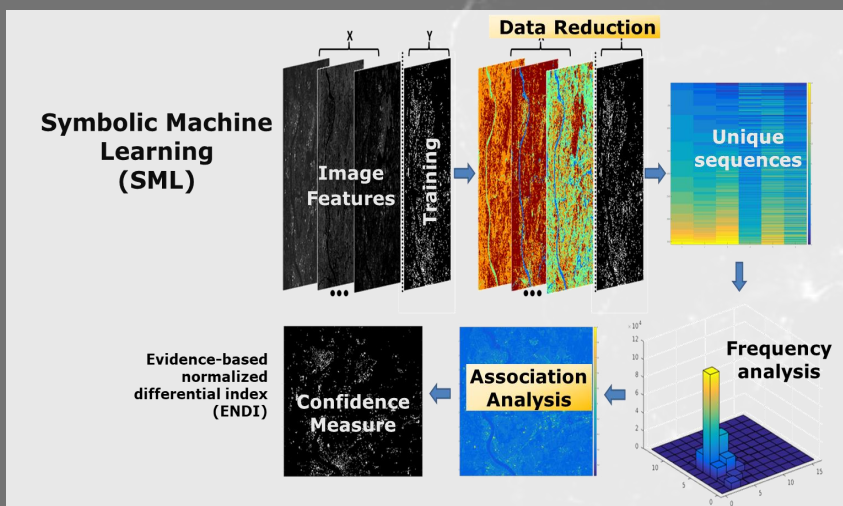


32 808 scenes

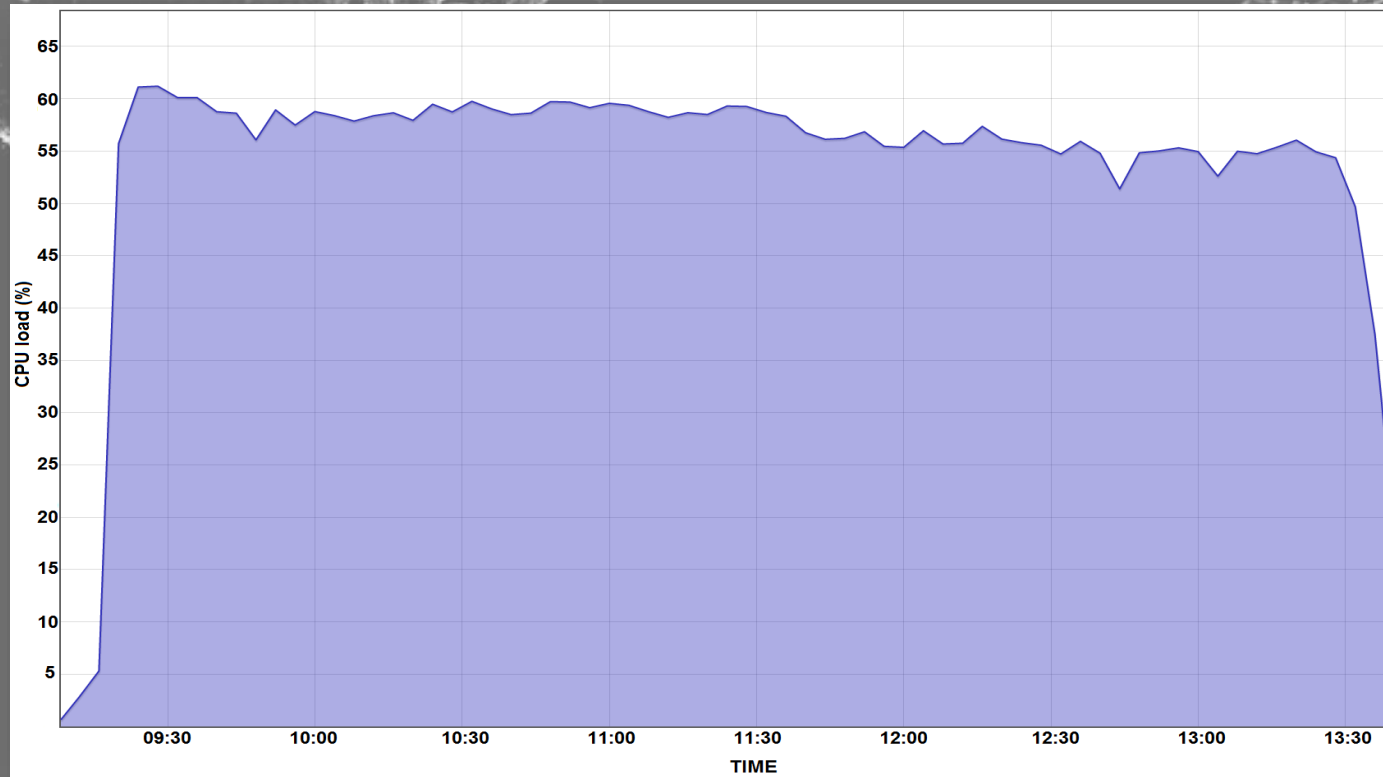
Processing of Landsat imagery on the JEODPP platform



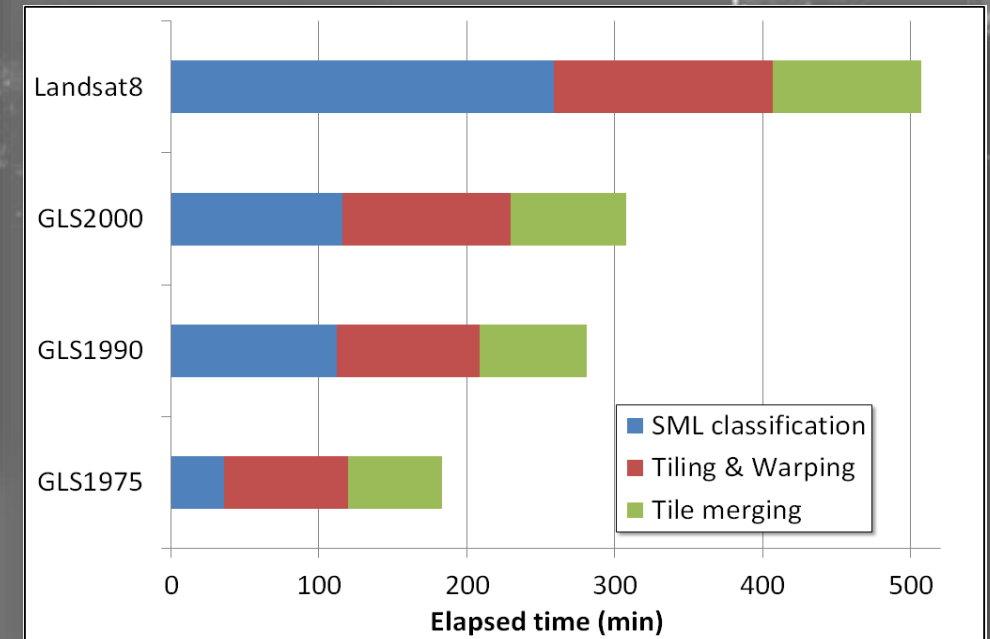
	Landsat collections
Input number of scenes	32 808
Input volume	23 TB
Output volume	14 TB
Processing time	22 h
Concurrent jobs	610 (SML) & 840 (warping, tiling and merging)



JEODPP CPU load during the processing of 9440 Landsat8 scenes



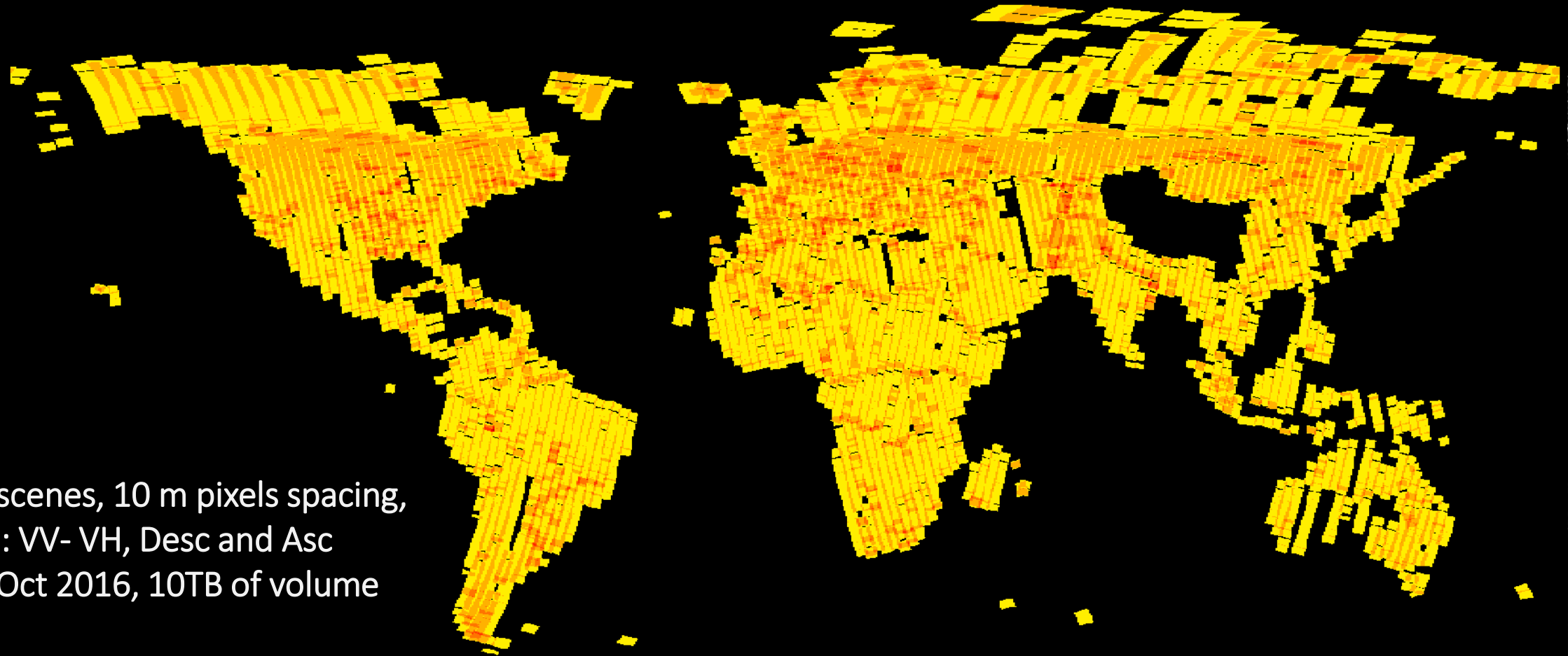
Total elapsed time Landsat processing (22h)



Sentinel-1 GHSL



Global coverage of Sentinel- 1



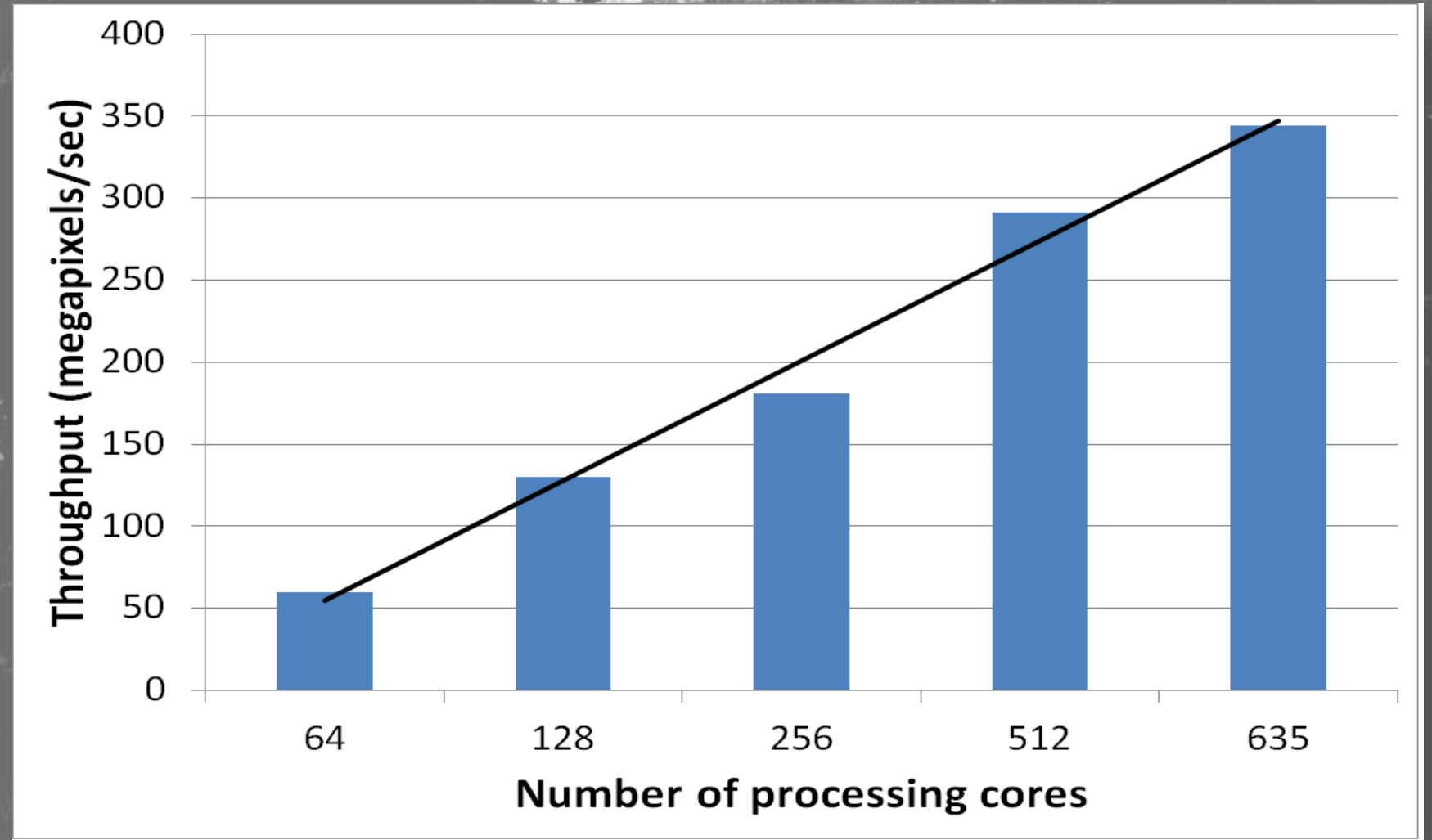
7000- GRD scenes, 10 m pixels spacing,
Polarization: VV- VH, Desc and Asc
Dec 2015 - Oct 2016, 10TB of volume



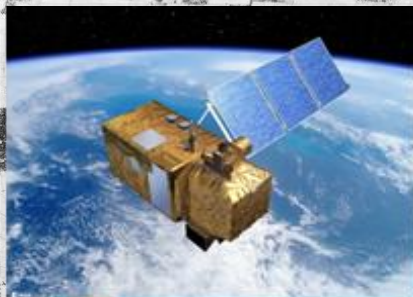
Processing of Sentinel-1 and Landsat imagery on the JEODPP platform

	Sentinel-1	Landsat collections
Input number of scenes	7000	32 808
Input volume	10 TB	23 TB
Output volume	23 TB	14 TB
Processing time	18 h	22 h
Concurrent jobs	240	610 (SML) & 840 (warping, tiling and merging)

Scalability of the SML classification algorithm while running on JEODPP



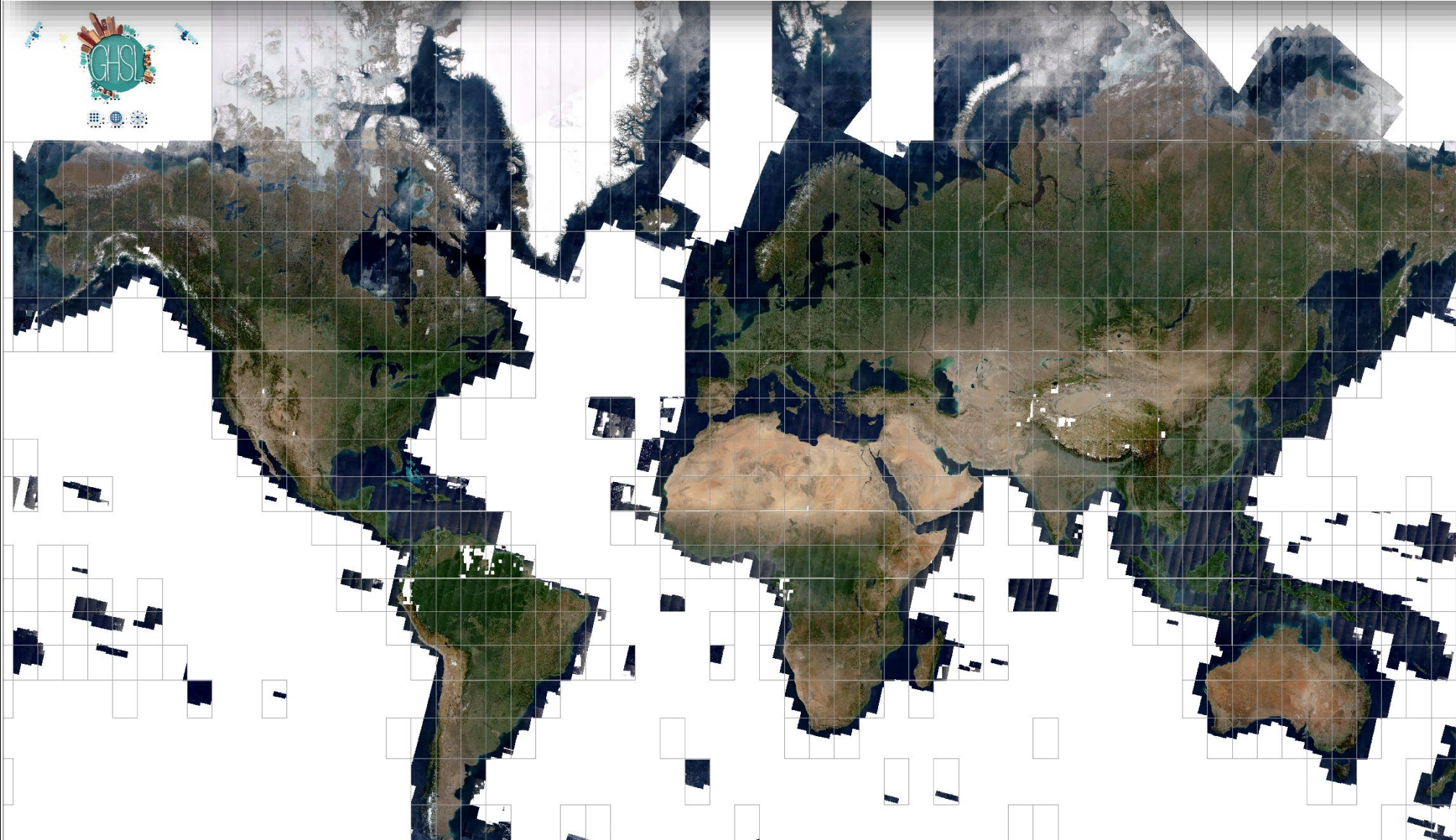
Sentinel-2 GHSL



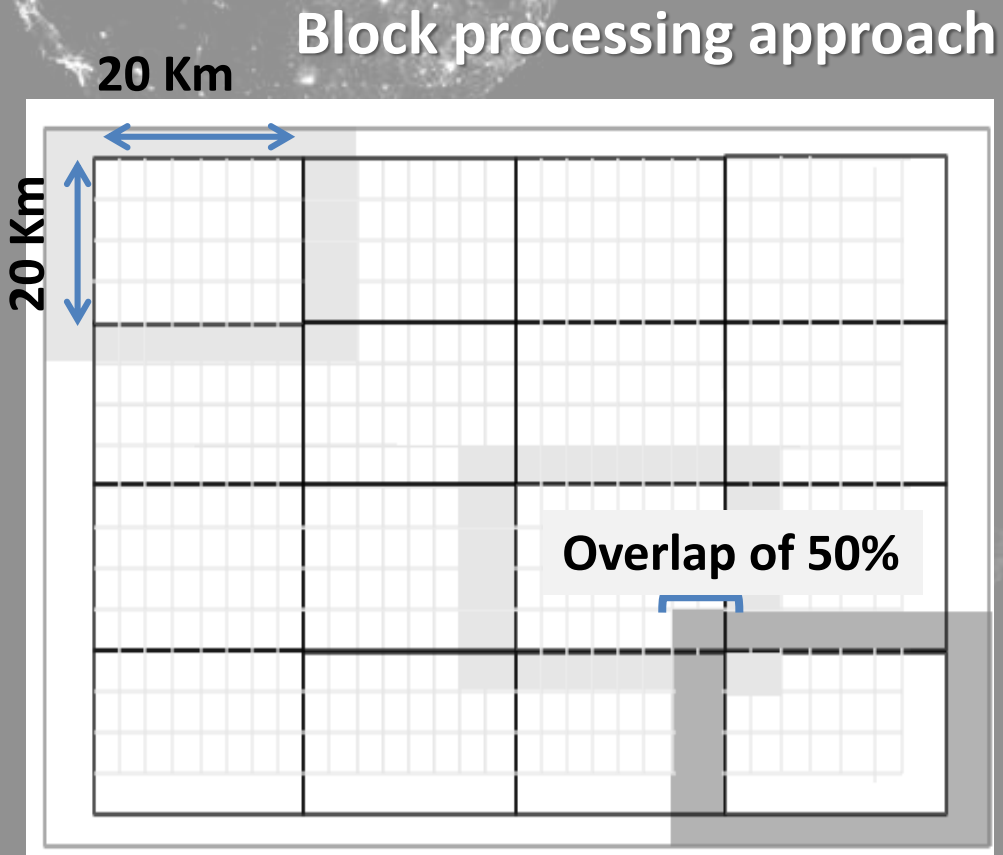
Global coverage of Sentinel- 2



10 Terabytes of imagery computed in and extracted from Google Earth Engine by UTM grid zone
Four bands: B2 (Blue), B3 (Green), B4 (Red), B8 (VNIR)
Spatial resolution: 10 meters - Level 1C

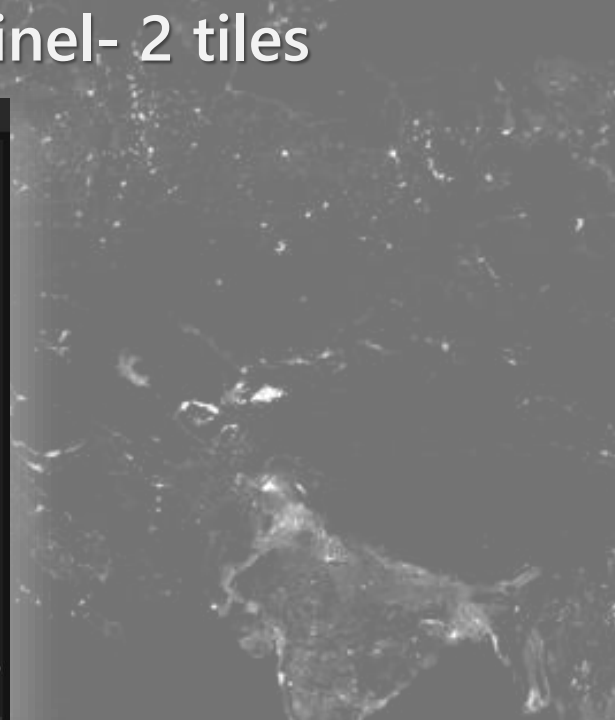
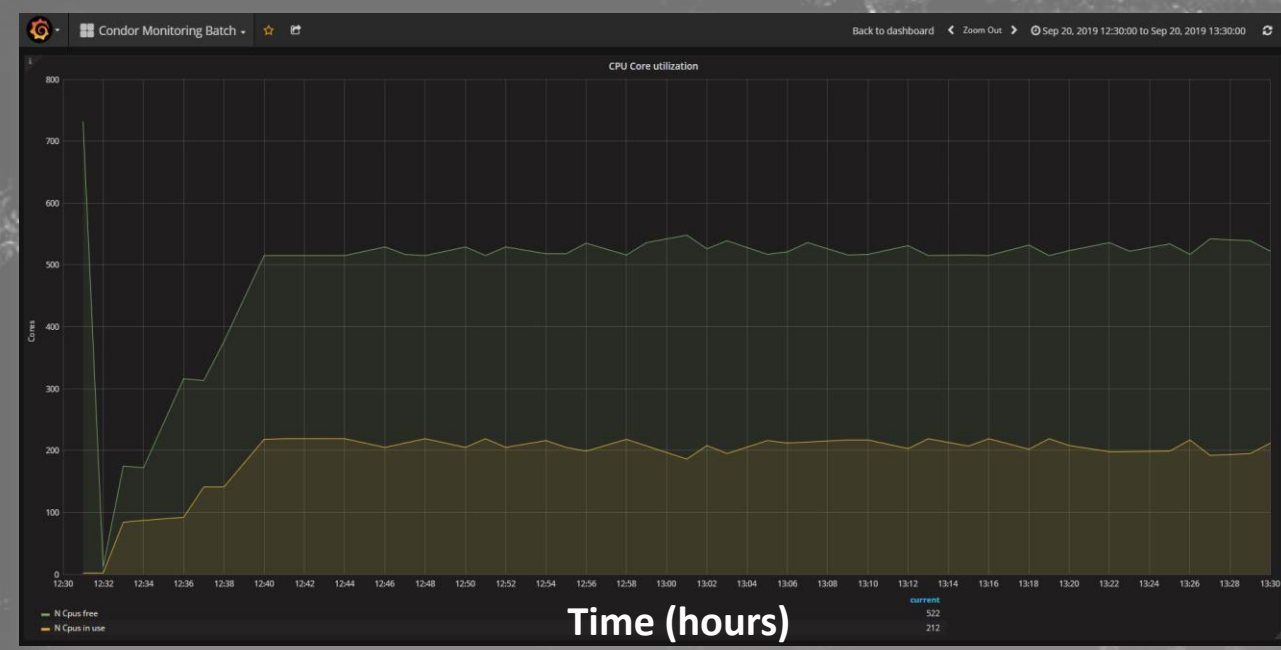


Processing of Sentinel-2 Global Mosaic at 10 meters (4 Bands)

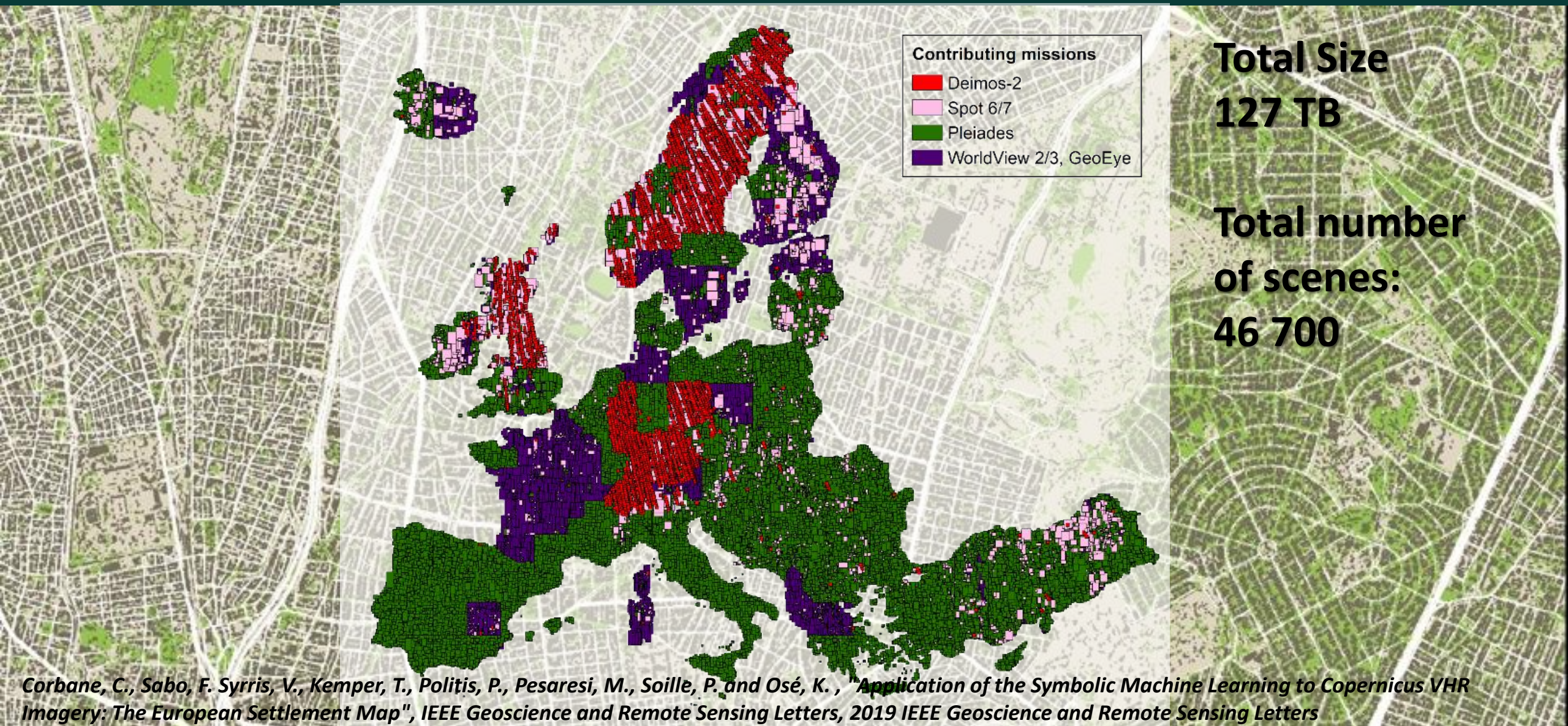


	Sentinel-2
Input tile/scenes	610 UTM grid zones divided into 20x20 tiles with 50% overlap => $610 \times 400 =$ 244 000 tiles
Input Spatial resolution	10 meters
Output volume	10 TB
Processing time	~3 days for processing at global scale
Concurrent jobs	80-200
Number of CPUs	1
RAM requirements per job	25-40 GB
Storage space on JEODPP storage disk	1 TB (including warping to a single projection)

JEODPP CPU load and running jobs during the processing of Sentinel-2 tiles over 1 Hour



European Settlement Map



Processing of Copernicus VHR data

European scale (Copernicus VHR for 2015)

The image shows a Linux desktop environment with a terminal window and a file manager window. The terminal window displays the execution of a Condor job script named `submit_sml_proc.sh`. The file manager window shows the contents of the `/eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts` directory.

Terminal Output:

```
sabofil@jeodpp-terminal-151p-04:/eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts$ condor_q ghs1proc
-- Schedd: s-jrciprjeop161p.cidsn.jrc.it : <139.191.240
OWNER  BATCH_NAME  SUBMITTED  DONE  RUN  IDLE
ghs1proc ID: 33767    2/6  17:01   959   336  45387

Total for query: 45723 jobs; 0 completed, 0 removed, 45723 held, 0 running, 0 suspended
Total for all users: 46927 jobs; 0 completed, 0 removed, 46927 held, 0 running, 0 suspended

sabofil@jeodpp-terminal-151p-04:/eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts$ condor_q ghs1proc
-- Schedd: s-jrciprjeop161p.cidsn.jrc.it : <139.191.240
OWNER  BATCH_NAME  SUBMITTED  DONE  RUN  IDLE
ghs1proc ID: 33767    2/6  17:01   1017   336  45329

Total for query: 45665 jobs; 0 completed, 0 removed, 45665 held, 0 running, 0 suspended
Total for all users: 46869 jobs; 0 completed, 0 removed, 46869 held, 0 running, 0 suspended

sabofil@jeodpp-terminal-151p-04:/eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts$ condor_q ghs1proc
-- Schedd: s-jrciprjeop161p.cidsn.jrc.it : <139.191.240.161:96187... @ 02/07/19 14:19:12
OWNER  BATCH_NAME  SUBMITTED  DONE  RUN  IDLE  HOLD  TOTAL  JOB_IDS

Total for query: 0 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
Total for all users: 6731 jobs; 0 completed, 0 removed, 4397 idle, 484 running, 1850 held, 0 suspended

sabofil@jeodpp-terminal-151p-04:/eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts$ condor submit -disable submit_sml_proc.sh
```

File Manager Content:

```
submit_sml_proc.sh
universe = docker
docker_image = jeoreg.cidsn.jrc.it:5000/jeodpp-htcondor/e1_ghsl_centos69_gdal21_py27_mcrb2016:1.0
+owner = "ghslproc"
executable = /eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/condor_scripts/start_sml_proc.sh
arguments = $(ClusterID) $(ProcID) /eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/jobs/Finland/finland_2.txt
#requirements = job_type == "dev"
request_memory = 40GB
request_cpus = 4

output = /eos/jeodpp/htcondor/processing_logs/GHSL/temp/job$(ClusterID)
error = /eos/jeodpp/htcondor/processing_logs/GHSL/temp/job$(ClusterID)
log = /eos/jeodpp/htcondor/processing_logs/GHSL/temp/job$(ClusterID)

queue 6
```

File Manager Table:

Size	Type	Date Modified
0 bytes	plain text document	10/10/2018
698 bytes	shell script	Tuesday
698 bytes	shell script	Yesterday
761 bytes	shell script	Today
2.2 kB	shell script	Tuesday

Terminal Script Content:

```
#!/bin/sh
#built the file output name
FNAME=`echo "job-"$CLUSTER_"$N_LINE".out"`
echo $FNAME

ls -la /scratch2/ghsl/

# export ALGDIR=/usr/local/share/GHSL
# mkdir -p /usr/local/share/GHSL
export ALGDIR=/tmp
# mkdir -p /tmp

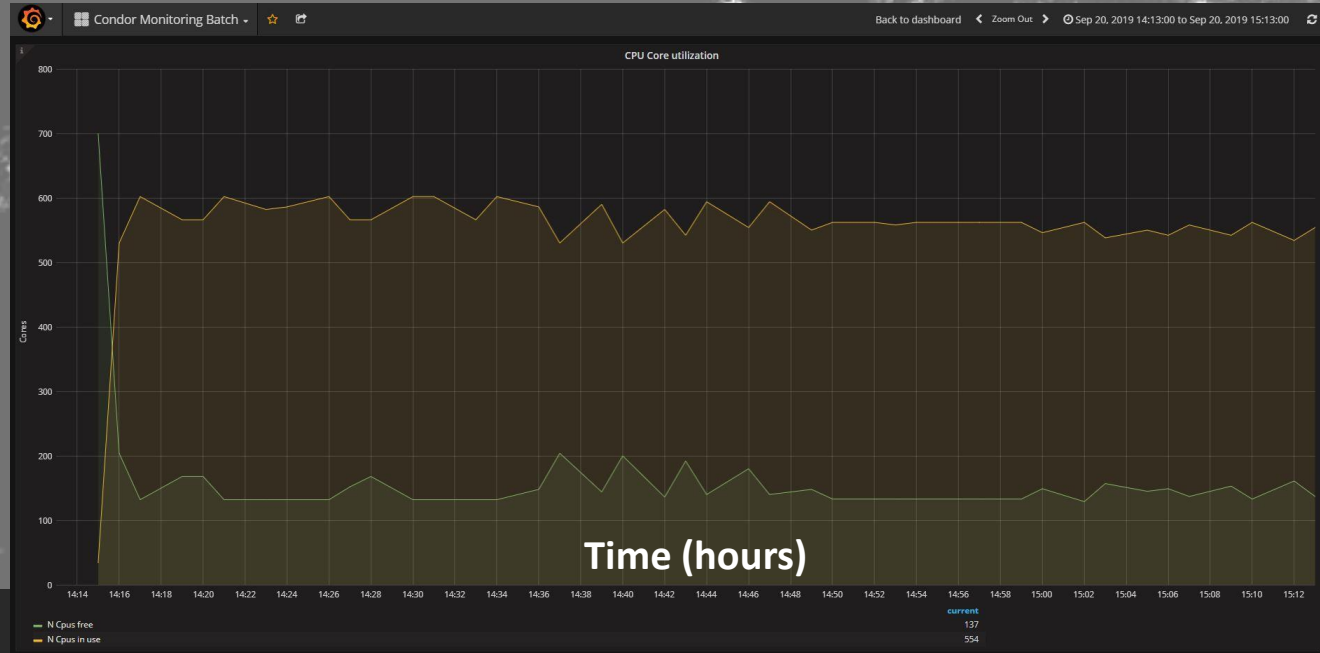
export MCRROOT=/usr/local/MATLAB/MATLAB_Runtime/v90
export GDAL_DATA=${ALGDIR}/libs/gdal-data

#create folders and copy files
cp -r /eos/jeodpp/data/projects/GHSL/ESM_2018/scripts/Batch_proc/matlab_executables/2018_ESM_Scandinavia/* $ALGDIR

cd $ALGDIR
nfname=`basename $INPUT_FOLDER`
nfname="${nfname}/."
echo $nfname
```

0 km

JEODPP CPU load and running jobs during the processing of VHR scenes over 1 Hour



Sep 20, 2019 14:13:00 to Sep 20, 2019 15:13:00



Lessons Learnt from processing of Big Earth Data

- **Autonomy in the execution of the workflows**
- **Flexibility and transparency in configuring the job requirements**
 - **Operating System**
 - **Number of CPUs**
 - **Memory**
 - **Specific machines**
- **Scalability of the processes crucial for continuous updates of Human Settlements Maps and Datasets**

THANK YOU

Dr. Christina Corbane
Christina.corban@ec.europa.eu



Tools



Global Human Settlement Layer (GHSL),
Joint Research Centre (JRC):
<http://ghsl.jrc.ec.europa.eu>

GEO GROUP ON
ARTH OBSERVATIONS

GHSL Supports
the 2030 Development Agenda

