

# Big Data Analytics for agricultural monitoring using Copernicus Sentinels and EU open data sets pilot

**Principal Investigators:** Guido Lemoine (Guido.LEMOINE@ec.europa.eu)

**Shepherd:** [Enol Fernandez del Castillo](#)

**Entry in the community requirement database:** [Big Data Analytics for agricultural monitoring using Copernicus Sentinels and EU open data sets pilot \(EAP\)](#)

## About the pilot

The project scope is, in principle, open to any scientific, public and private data user who may derive added value in the agricultural use domain, on the condition that feedback on required and demonstrated functionalities and performance is provided. The key aspect in the early adopter demonstrator is to show how federated EOSC resources can facilitate a range of Sentinel data applications across agricultural user domains. This extends the Copernicus DIAS concept, aimed at business users, to scientific and public users, by ensuring interoperability between EOSC resource providers and exposing the Copernicus high resolution Sentinel data archive with standardised processing services through tested standard interfaces.

## Team

Participant	Role	Name and Surname
JRC	Community PI	Guido Lemoine
EGI.eu	Shepherd	<a href="#">Enol Fernandez del Castillo</a>
CESNET	EGI Cloud Compute Provider	<a href="#">Zdenek Sustar</a>
CloudFerro	Cloud Compute Provider	<a href="#">Marcin Gil</a>
EODC	Cloud Compute Provider	<a href="mailto:Christoph.Reimer@eodc.eu">Christoph.Reimer@eodc.eu</a>

## Technical Plan

Goal(s) to be achieved after 1 year	<ol style="list-style-type: none"><li>1. Full Sentinel-1 and -2 territorial coverage of Netherlands and North Rhine-Westphalia and the application of machine learning techniques to fulfil CAP monitoring requirements.</li><li>2. Extension of the Groen Monitor (<a href="http://groenmonitor.nl">groenmonitor.nl</a>) with Sentinel-1 time series for crop phenology monitoring and specific crop management practices (e.g. sowing, harvest, soil tillage) to complete the existing Sentinel-2 based services.</li><li>3. Correlation of phenometrics derived from Sentinel-2 with other sources, such as phenology networks, coarser spatial resolution satellite products, temperature-driven phenological models, and ground and/or volunteered phenological observations.</li><li>4. Monitoring inter-regional variability in derived indicators for cash crops, establishing robust statistical estimators for intra-field and intra-region comparison of ready-to-use variables derived from Sentinel time series, with a focus on sugarbeet and potatoes. Using machine learning algorithms to correlate predicted trends to extensive field observations and disseminate results to service users.</li></ol>
<b>Roadmap/work-plan</b>	
Work planned for Q1	<ol style="list-style-type: none"><li>1. Infrastructure resource provisioning:<ol style="list-style-type: none"><li>a. 1 Database server VM with 4 vCPUs, 16GB RAM, 1TB local storage</li><li>b. 2 high-end VMs: 16 vCPUs / VM, 64GB RAM / VM, 100GB storage / VM</li><li>c. Additional 10TB S3 type block storage</li></ol></li><li>2. Completion of S1 and S2 time series for 2019, workshop with partners on data access, Jupyter notebook development.</li><li>3. Demonstration of data use in CAP policy context, using 2019 reference data</li></ol>
Work planned for Q2	<ol style="list-style-type: none"><li>1. Continued support to partners on data analytics, integration of specific processing needs in compute infrastructure</li><li>2. Development of alternative data formats for interactive analysis and visualization</li><li>3. Initiation of 2020 data processing</li><li>4. Technical documents and tutorials on the EOSC enabled analysis workflow + release of open source components</li></ol>

Work planned for Q3	<ol style="list-style-type: none"> <li>1. Continued processing for 2020 + data extraction, monitoring tools to review operational workflow</li> <li>2. Support to partner experiments with data sets with additional data processing routines</li> <li>3. Exposure of EAP project results in workshops</li> </ol>
Work planned for Q4	<ol style="list-style-type: none"> <li>1. Completion of 2020 data processing</li> <li>2. Reporting on technical development and data use demonstrations</li> <li>3. Final demonstration materials, tutorials, code base</li> <li>4. Review of EAP experience and recommendations for (continued) operational scale up.</li> </ol>

## EOSC services and providers

EOSC Service	Provider	Resources to be committed	Status
	EODC	8 vCPUs / 32GB of RAM,/ 100GB HDD	Resources allocated and available
<a href="#">CloudFerro Infrastructure</a>	CloudFerro	24 vCPUs / 64GB of RAM / 100GB HDD	Resources allocated and available
<a href="#">EGI Cloud Compute</a> and <a href="#">EGI Online Storage</a>	CESNET	16 vCPUs / 64GB of RAM / 100GB HDD	Resources allocated and available
		32 vCPUS / 128 GB of RAM / 200 GB HDD	Resources requested, waiting for allocation
		Database server: 4 CPUs / 16GB / 1TB HDD	Resources allocated and available
		20 TB object storage	Resources requested, waiting for allocation