



# BDA needs for the Earth Observation Science Community

EGI GOSC workshop session 2 - 3 November 2020

*Guido Lemoine, European Commission,  
Joint Research Centre*

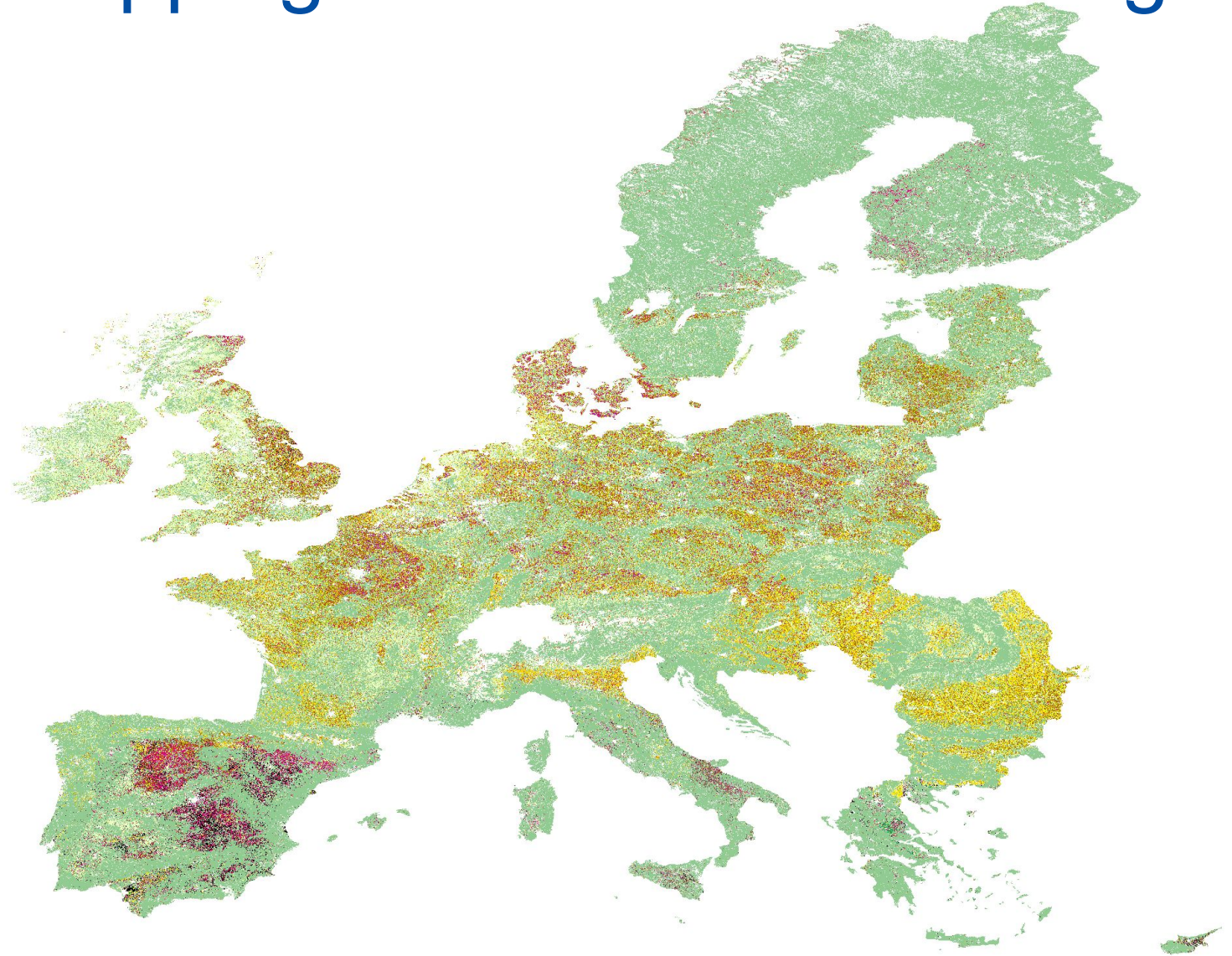
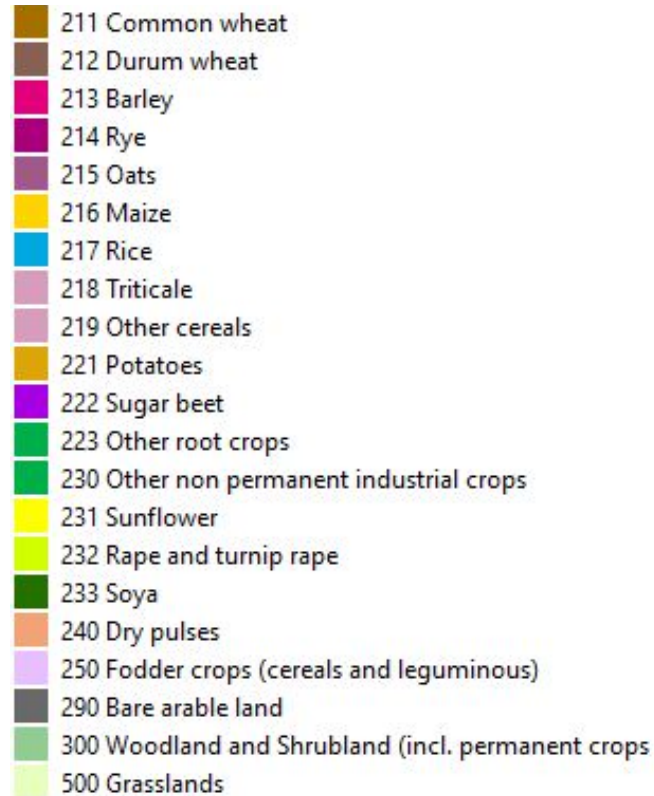
# Scope

- Earth Observation (EO) science is undergoing a **paradigm change**
- Open access to high quality EO data from EU Copernicus Sentinels, US Landsat (soon NISAR), Japanese ALOS, Chinese Gaofen, etc.
- At data rates of multiple TBs/day, Petabyte-scale annual archives
- **Big Data Analytics** (BDA) required, based on modern machine and deep learning techniques, coupled with more complex physical models, at higher spatial, temporal and hybrid radiometric resolutions.
- Integrating richer reference and validation data
- Enabling discovery **at scale** in natural and social science domains beyond the traditional remote sensing and spatial analysis disciplines

# Core requirements

- EO science use patterns are diverse, from local to global monitoring, but increasingly towards global applicability (e.g. SDGs)
- The key requirement is permanent open access to full archives (**data curation**)
- Sensor data as **Application Ready Data** (not the case for Sentinel-1)
- If full archive is distributed, **high connectivity** and **smart caching** and **prediction** is needed
- Global upscaling implies use of massively parallel cloud resources
- Processing needs imply use of special hardware solutions (e.g. GPU)
- Distributed processing prescribes Open Source solutions

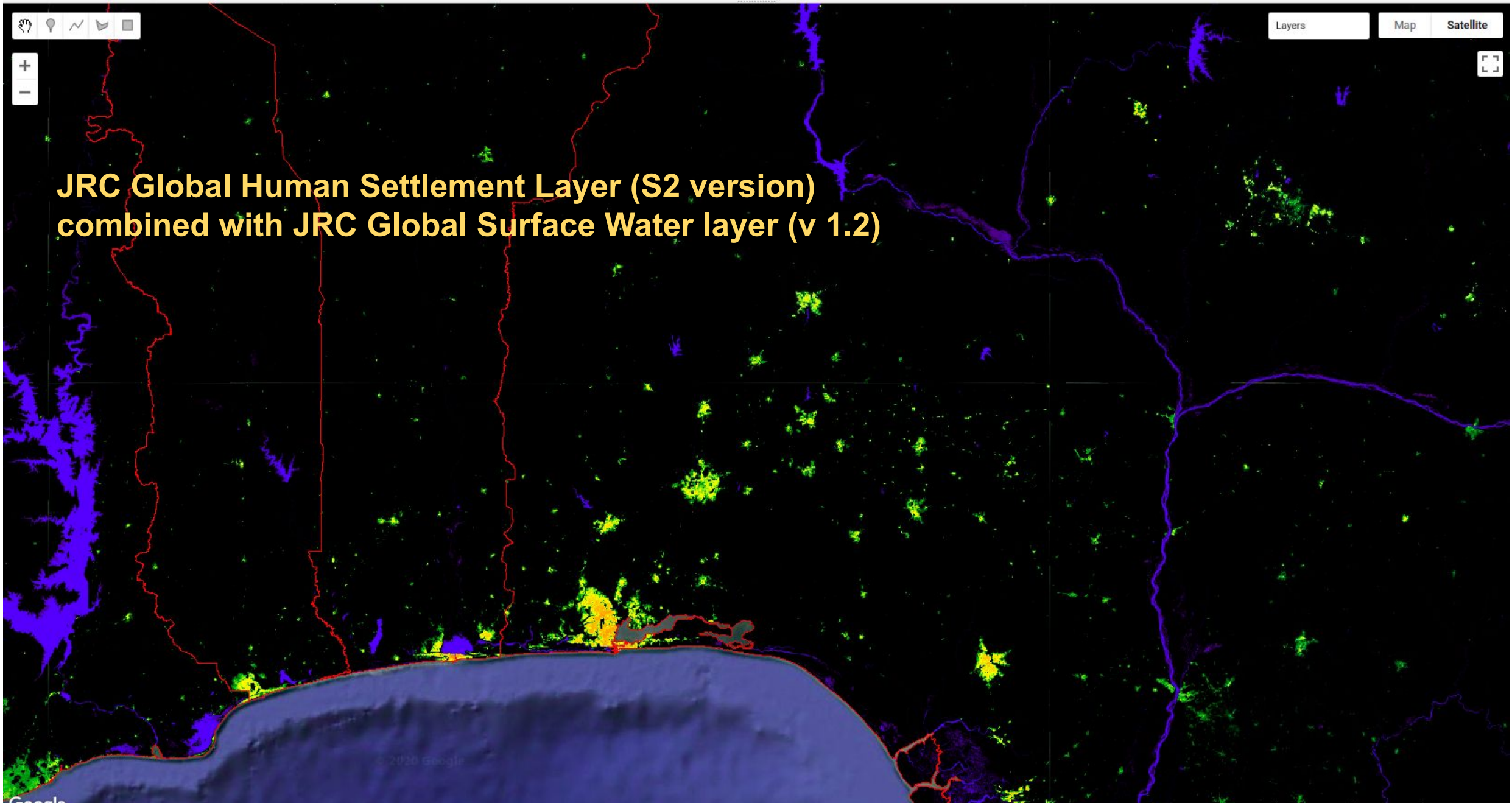
# Setting a European benchmark for continental scale crop type mapping with machine learning



*EU 2018 crop map at 10m pixel spacing. Masked with non-vegetation Corine Land Cover classes.*



**JRC Global Human Settlement Layer (S2 version)  
combined with JRC Global Surface Water layer (v 1.2)**



# Status

- Google Earth Engine is the *de facto* standard
- Not an open cloud infrastructure, but free for science use
- Especially the [interface to] massively parallel application of rich data analytics is unmatched by other cloud infrastructures.
  
- EU Copernicus DIAS offer IaaS closely coupled to a multiple Petabytes S3 archive (publicly financed until end of 2021)
- Some DIAS instances and smaller equivalents are federated in the European Open Science Cloud
- Familiarity and technical expertise to scale EO science use on open cloud solutions is still scarce.

# Outlook (1)

- In Europe the **Digital Europe** agenda is shaping up as the key initiative for a federated European Open Science Cloud
- **Destination Earth** has specific relevance for the EO science domain
- Extends to “European Data Spaces” to unlock thematic potential
- Scale **DOES** matter. Current open cloud landscape very fragmented, e.g. per application domain, national initiatives.
- Consolidation through coordination of funding streams (project compute requirements as in-kind open science cloud credits)

# Outlook (2)

- Europe has state-of-the-art Open Source geospatial expertise
- Increasingly in applied machine/deep learning, with rich reference data
- More efforts needed to promote cloud solutions in EO research
  
- Portable/Scalable to other continental open science cloud initiatives
- To accommodate additional open data streams, globally distributed archives, unlock validation and reference data, stimulate Open Science
- Little time left to be ready for upcoming open data streams, novel commercial data access mechanisms



# Thank you

[guido.lemoine@ec.europa.eu](mailto:guido.lemoine@ec.europa.eu)