A Hybrid Architecture for Tactical and Strategic Precision Agriculture

<u>Enrico Gallinucci</u>, Matteo Golfarelli, Stefano Rizzi

{enrico.gallinucci, matteo.golfarelli, stefano.rizzi} @ unibo.it

BUSINESS INTELLIGENCE GROUP

Introduction

The world of agriculture has become a major producer/consumer of data

- Precision agriculture (PA): farming management based on a DSS

Scientific literature poorly addresses the definition of a comprehensive and integrated architecture for precision agriculture

- Free and commercial data services provide high-value data
  - E.g., irrigation advices, vegetation indices
- Mostly based on web applications to deliver data services
- Strongly differ in:
  - The way data are stored, processed, and made available
  - The type of data provided
  - The professional figures and services they are oriented to

Introduction

In this paper we propose an innovative architecture for handling agricultural data in an integrated fashion

- Oriented to data analysis and inspired by Business Intelligence (BI) 2.0
- Hybrid: couples traditional and big data technologies
  - Integrate heterogeneous data, at different levels of detail, from owned and open sources

Enables different levels of tactical and strategic services

- Tactical: **detailed information** from a **limited area** within a **restricted time-span**
- Strategic: **aggregated data** from **broader areas** spanning on **long time intervals**
- Separated repositories and schemata are required
- Geo-referencing exploited as the basis for painless integration

Related work

Most efforts focused on applying **ML to ad-hoc agricultural datasets**

- Marginal focus on data management architecture and platform's features

Chen et al. (2015)

- Focus on data collection and tactical level
- **No strategic level/OLAP functionalities**
- **No validation on big data**

Free/commercial web-services

- Global Land Cover, CropScape, Mundialis, Moses, Earth Observation Data Services
- **No OLAP-like functionalities**

The project

Architecture developed in the context of the Mo.Re.Farming project

- Provide a Decision Support System for agricultural technicians in the Emilia-Romagna region (Italy)
- Enable analyses related to the use of water and chemical resources
  - In terms of optimization and environmental impact

Our goal: verify the feasibility of a data integration approach

- Deliver precision agriculture services to a plethora of different stakeholders
- The more effectively integrated data sources,
  the more delivered services and supported stakeholder

Data sources

Satellite images

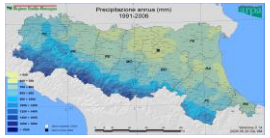
- Source: ESA (European Space Agency)
- Taken from Sentinel-2 satellites
- Each image is a granule in USA's National Grid
  - www.fgdc.gov/usng
  - About 1GB each
  - Emilia-Romagna region covered by 7 granules
- Designed to support vegetation, land cover, and environmental monitoring
- Provided in the L-1C format
  - Affected by solar light's reflection against atmosphere
- 3 satellite passages per week

Data sources

Field sensors

- Humidity sensor
  - Uses a waveguide faced to the soil surface
  - One value per hour
- Smart pheromone trap
  - Insects captured through an adhesive strip with pheromones
  - Smart camera counts and classifies insects every day

Data sources



Weather data

- Source: ARPAE (Regional agency for climate monitoring)
- Temperature and rainfall data from 858 sensors in Emilia-Romagna
- Updated daily

Crop & Rural land registers

- Sources: AGREA, CER (Regional agencies for agriculture)
- Vector layer with field's surfaces and classifications
- Updated yearly

Administrative boundaries

- Source: ISTAT (National institute of statistics)
- Vector layer municipal, provincial, and regional boundaries

A closer look

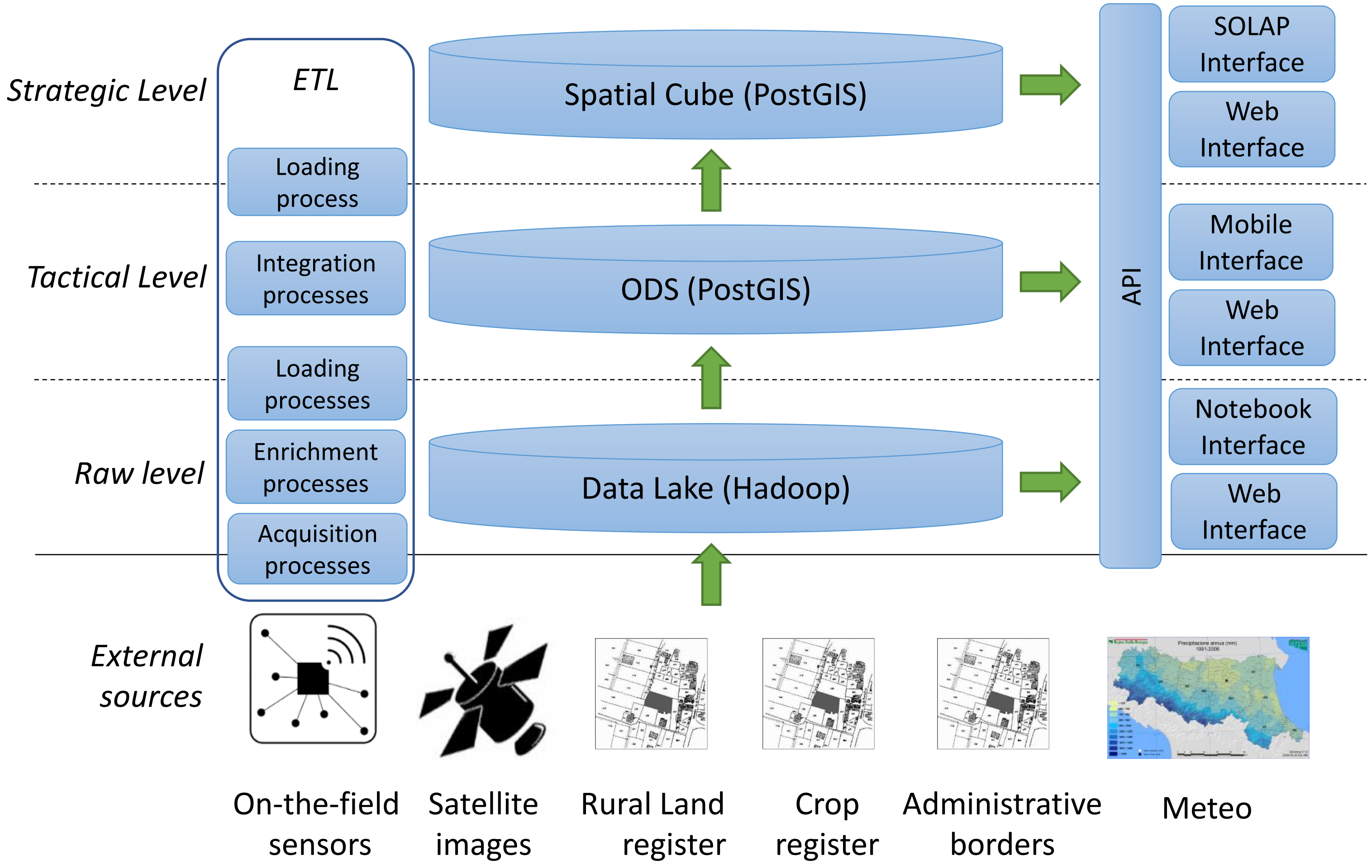
The architecture

- From the data lake to the spatial cube

Data processing

- Acquisition, enrichment, integration and loading

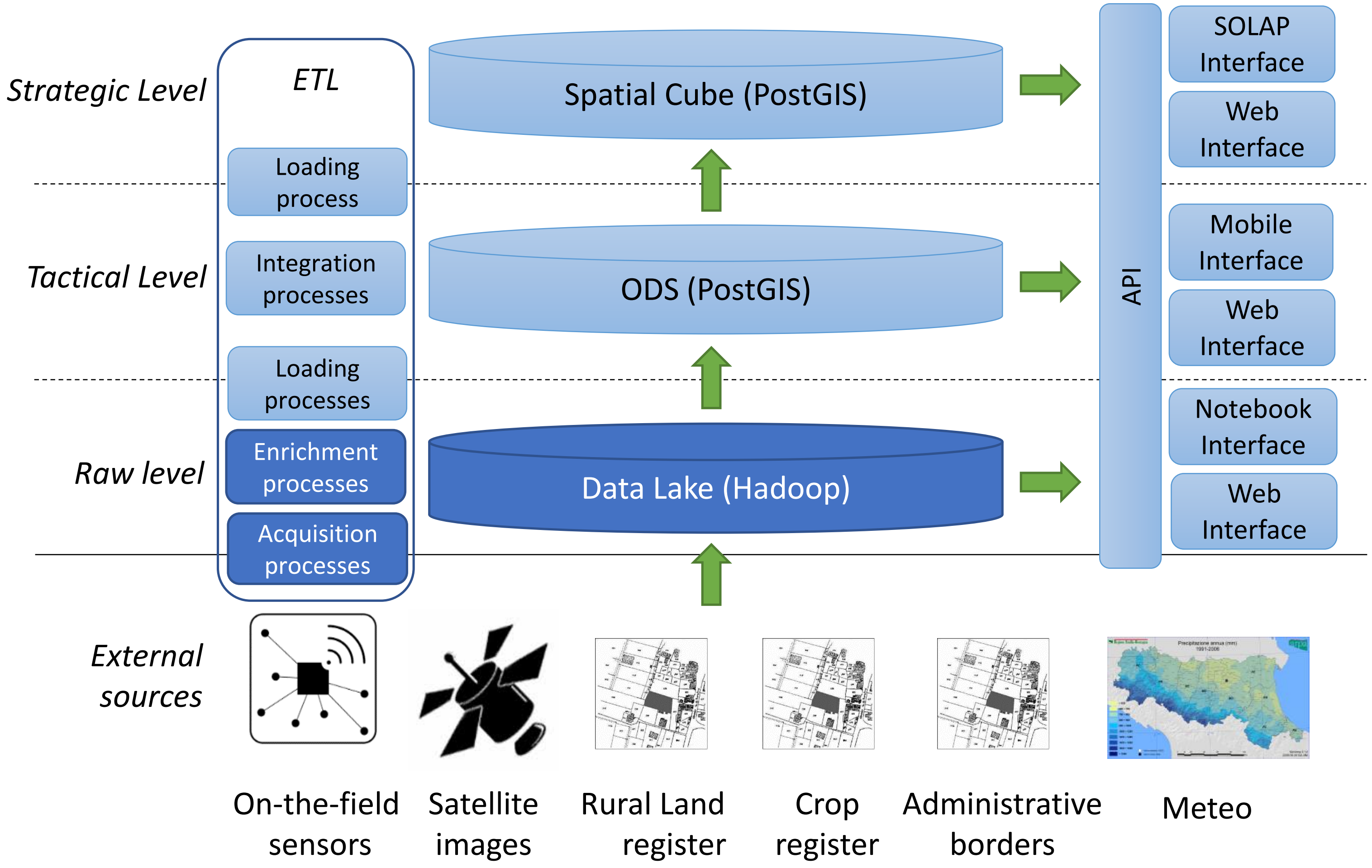
The architecture



The architecture

Data lake

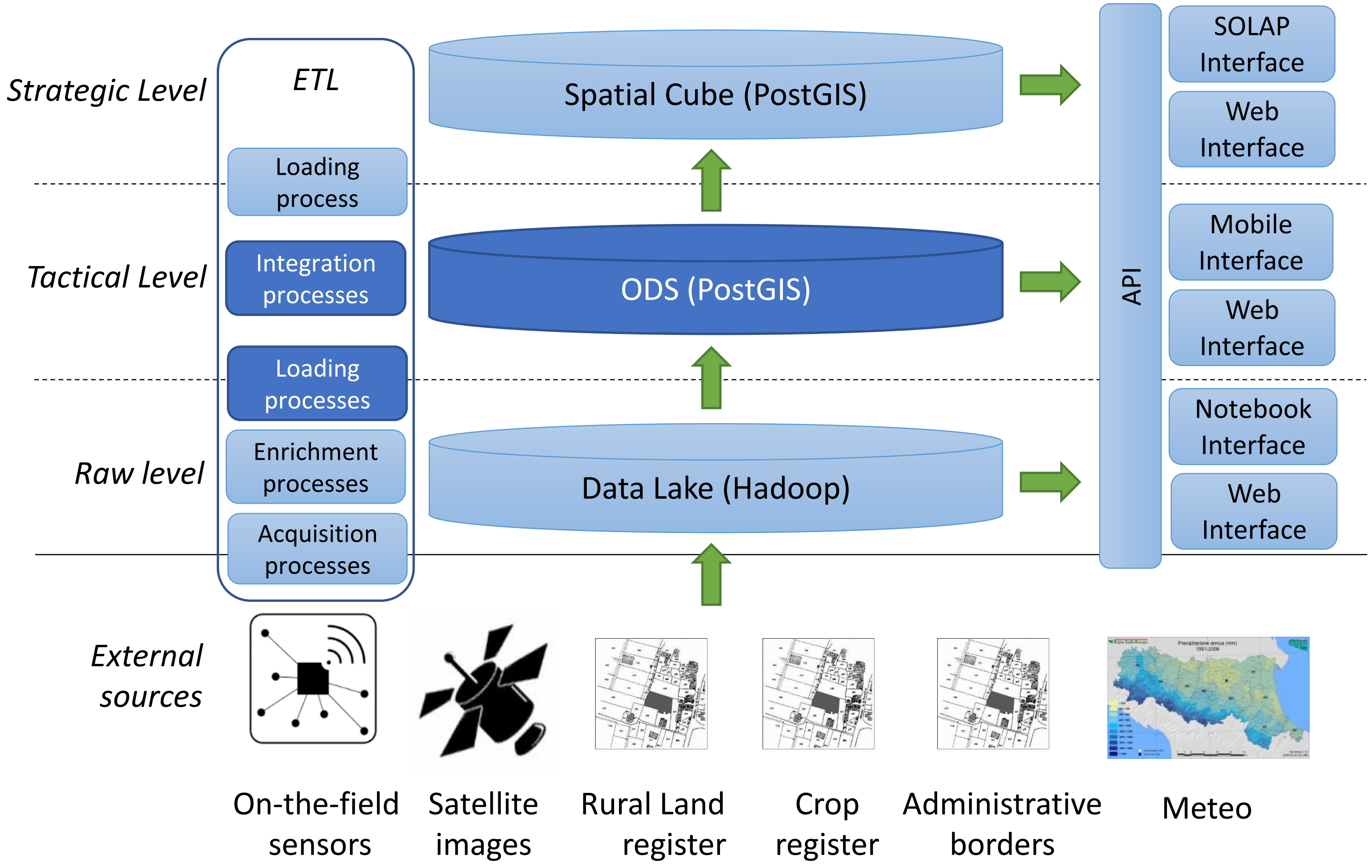
- Holds a vast amount of raw data in native format
- Adopts a multi-zone architecture
- No fixed schema



The architecture

Operational Data Store

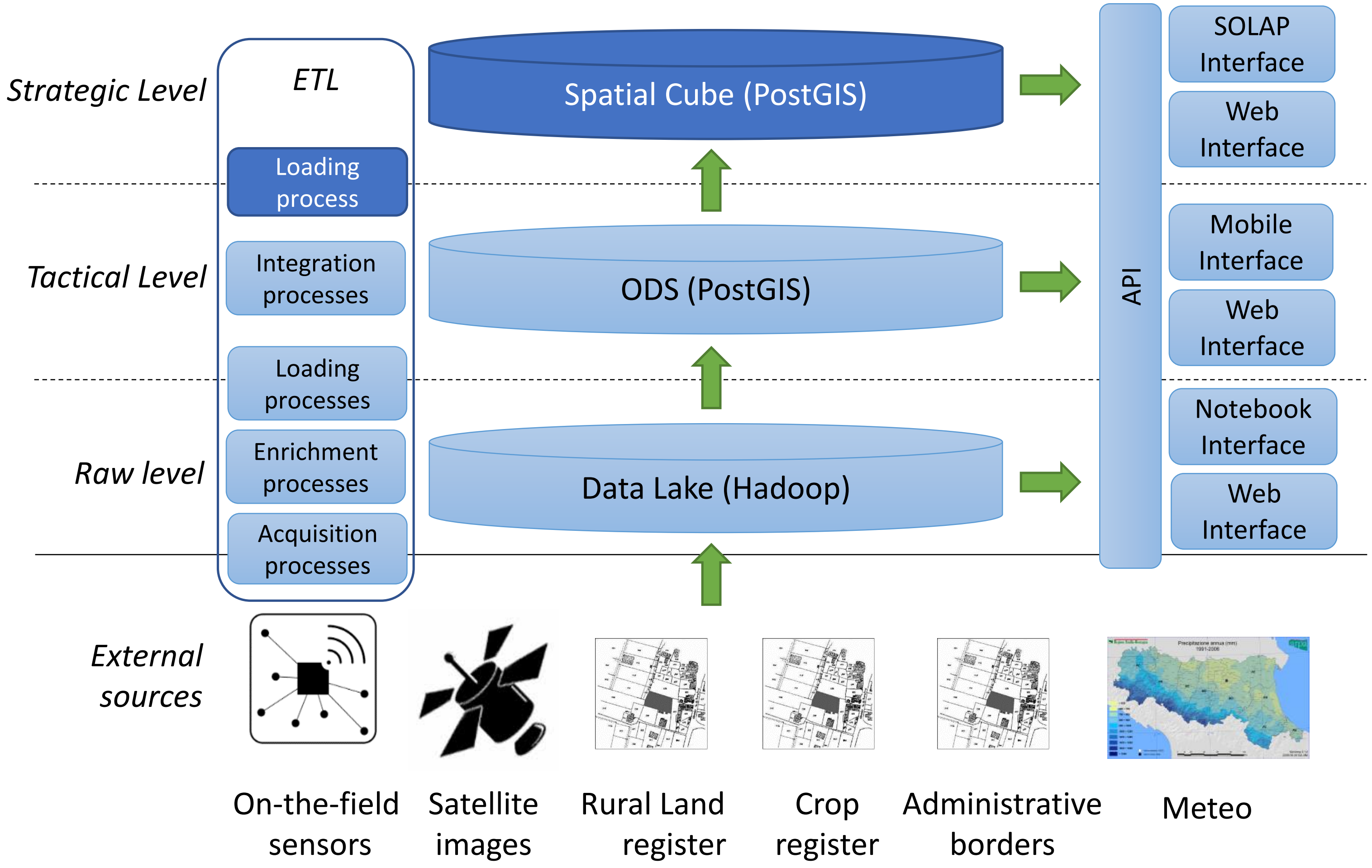
- Stores structured data at the finest level of detail
- In-depth analysis and monitoring
- Schema defined at design time



The architecture

Spatial cube

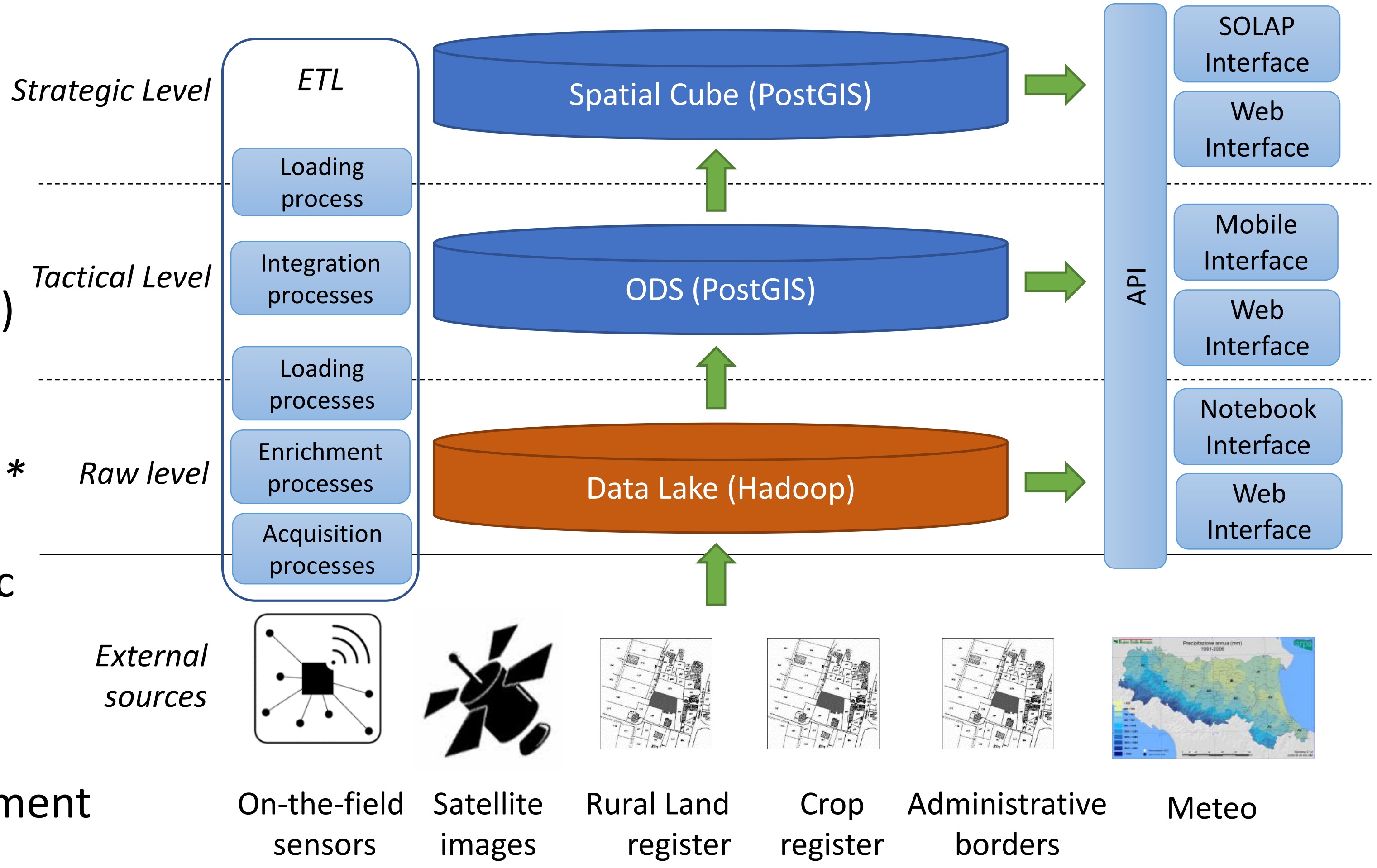
- Data stored in a multidimensional fashion
- Enables SOLAP (Spatial On-Line Analytical Processing)
- *Spatial slice* and *spatial drill* operators allow for aggregating measure values



The architecture

Hybrid architecture

- Data lake is Hadoop-based, upper levels rely on centralized PostGIS DBMSs
- Satellite images (i.e., rasters) need to be handled at the pixel level
- Analysis of *continuous fields** requires *map algebra manipulations*, i.e., algebraic operations on rasters (e.g., arithmetical, trigonometric)
- No Big Data GIS available at the time of project development



* Spatial phenomena that are perceived as having a value at each point in space and/or time

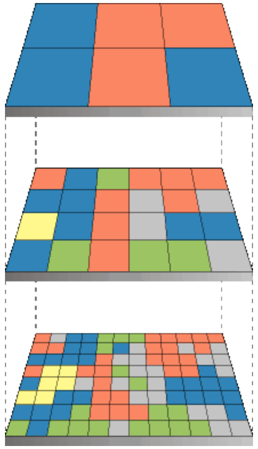
Data acquisition and enrichment

HDFS ensures system robustness and enables parallel processing

- 11-node cluster (4-core CPU, 32GB RAM, 6TB disk)

Satellite image acquisition and enrichment parallelized as bag-of-task:

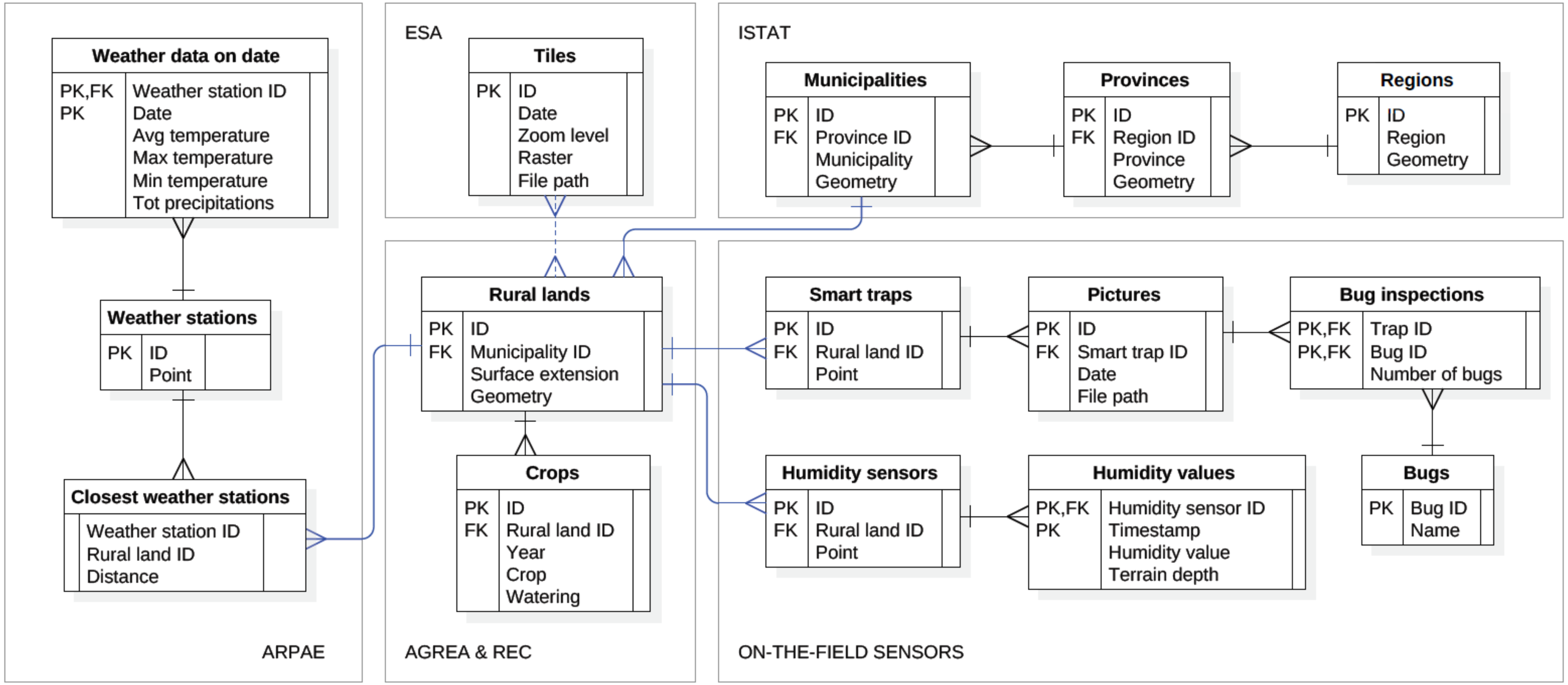
- Download through ESA's web services
- Atmospheric correction
  - Cleans solar lights' reflection on atmosphere
- Merging granules into a single GeoTIFF
- Raster pyramid creation
  - Enable visualization through a web interface
  - Split into tiles at several resolution levels
- Sequential ETL for other sources



Data integration

The central role is played by the Rural lands

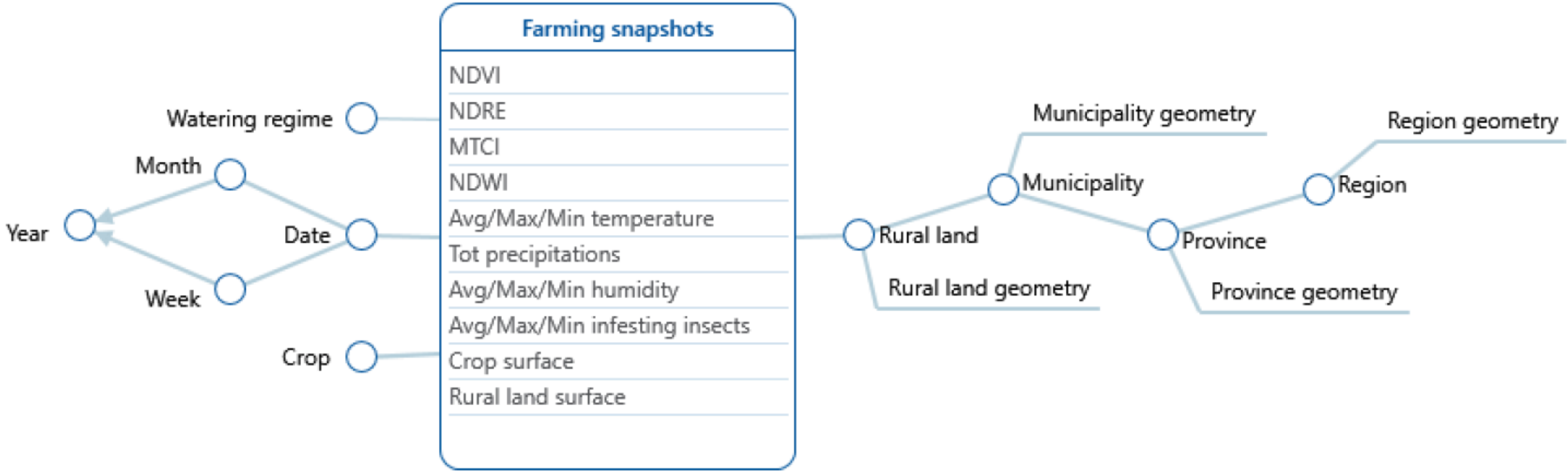
- Enabling the spatial integration of the different data sources



Data loading

Multidimensional schema obtained with a data-driven approach

- Validated by experts in the field of agricolture
- Events consist of **snapshots**, one for each satellite image
- Provide statistics for the crop of a given rural land in a given date



Performance

Spatial cube loading is the most expensive

- Due to spatial integration of rural land vectors and tile rasters to compute the vegetation indices
- Loading is incremental

Variations due to

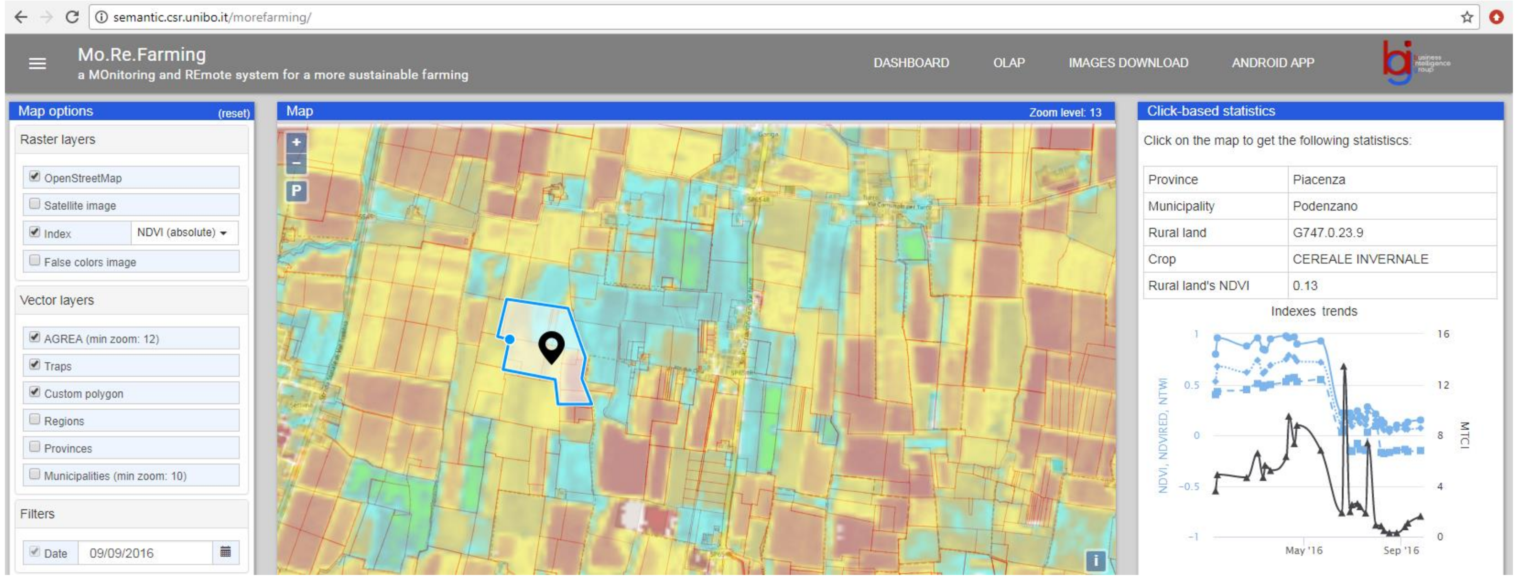
- Image size
- Number of granules per image (from 1 to 7)

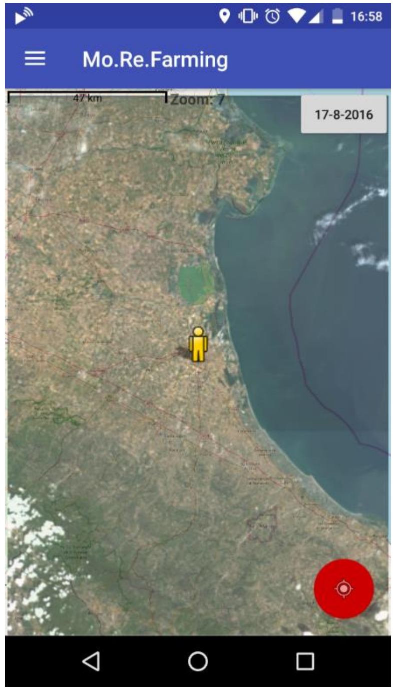
Process	Execution time	Granularity
Satellite image acquisition	2-5 minutes	Per granule
Atmospheric correction	25-40 minutes	Per granule
GeoTIFF creation	10-40 minutes	Per image
Pyramid creation	5-15 minutes	Per image
ODS loading	15-20 minutes	Per image
Spatial cube loading	3-8 hours	Per image

Tactical level interface

Ad-hoc dashboard: http://semantic.csr.unibo.it/morefarming

- Obtain statistics about a point, field, or custom area
- Locate smart sensors and visualize the generated data
- Also available as a mobile application

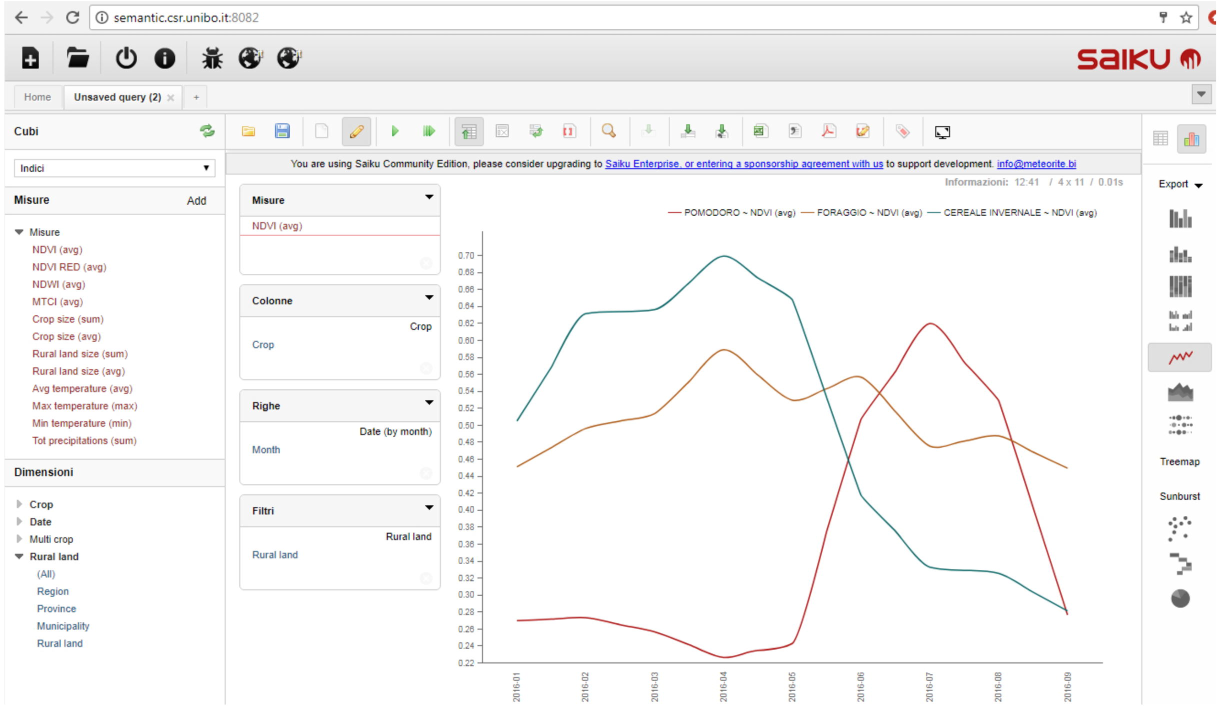




Strategic level interface

SOLAP solution implemented through Saiku

- Polygon drawn on the map used as a spatial filter in the query



Standard RESTful APIs: http://semantic.csr.unibo.it/morefarming/api

Conclusions

Technical considerations

- Computational needs and data quantities push towards big data solutions
  - The level of development of tools with spatio-temporal features is still limited
- Carrying out integration at the physical level has both pros and cons
  - + Late data reworking
  - + 360-degrees exploitation of data with no limitations due to different data owners
  - + Possibility of efficiently running complex analytics on heterogeneous data
  - - Integration tasks are computationally demanding; hard to be carried out on-the-fly
  - - Higher costs for storing and handling data
- New data sources and data collection approaches in the coming future
  - *Farmsourcing*: crowdsourcing applied to agriculture through IoT machinery and sensors

Conclusions

Domain-related considerations

- Creating integrated hubs of information is mandatory to deliver more effective information and services
  - Several solutions, models and services for PA
  - Work on a subset of the available data, providing complementary non-integrated info
- PA should be as open as possible to enable complementary data exchange and fruition
  - Need for a standard and machine-readable terminology
- Stakeholders have different skills, culture, and mindset
  - Technicians/managers favorably perceive a data-driven agriculture; most farmers do not
  - Overcome reluctance with easy-to-use analytics and killer applications
  - Mo.Re.Farming mobile interface strongly appreciated by in-field users

Up next: Agro.Big.Data.Science

New project to complete the picture

- Moving the platform to a full Big Data infrastructure
  - Exploit new software tool for continuous field analysis (i.e., GeoTrellis)
- Mobile application for remote feedback by farmers
  - Designed in collaboration with technicians
  - To be coupled with on-field sensor data
- A more intelligent platform
  - Smart data lake for automatic metadata extraction and lineage tracking
  - General-purpose visualization of geo-located data
- Decision support
  - Actively support the decision-making process beyond the mere integration of data

Thank you

Questions?

BUSINESS INTELLIGENCE GROUP