The e-Infrastructure for large data analytics in agriculture

F. Javier Nieto – ATOS Spain

Athens, 10th July 2019

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 777549
Outline

1. EUXDAT Context & Objectives
2. Global Picture
3. What Can EUXDAT Offer?
4. Overview of the Achievements
5. Challenges to be Addressed
EUXDAT Context (I)

http://www.euxdat.eu/
EUXDAT Context (II)

- Open Land Use Map Improvement
- Monitoring of Crop Status
- Delimiting Agro-Climatic Zones
- Looking for Climatic Patterns Changes
- Information Support for Field Use Recommendations
- Effective Utilization of Natural Resources

http://www.euxdat.eu/
Objectives & Context

“...EUXDAT will build up a Large Data Analytics-as-a-Service e-Infrastructure with several software layers supporting sustainable and productive agriculture...”

- Manage data storage and movement + Support heterogeneous data sources + configurable policies
- Adapt data processing tools for HPC + Users’ Portal with advanced features + Hybrid HPC&Cloud resources management
- Provide access to EUXDAT services + Pilots implementation
- Facilitate long-term sustainability + Collaboration (i.e. PRACE, EOSC)
Example: Delimiting agro-climatic zones

- Elevation as a factor influencing temperature (DEM)
- Slope orientation as a factor influencing temperature
- Hydrology as a factor influencing temperature (buffering)
Global Picture (II)

Scenario specific GUI

MapServer
WMS

API (Python Django)
- Field description
  (= polygons)
- Results
  Raster (healthiness map)

Database (PostGIS)
- Polygons
- Resulting statistics (JSON)

Olive monitoring
Python Script

Orchestrator
Cloudify

execute

Sentinel API

http://www.euxdat.eu/
Global Picture (III) - Interfacing with DIAS

* Image from Budapest ICT Proposers Day (11/2017) – Space WP

http://www.euxdat.eu/
What Can EUXDAT Offer?

- Management and analysis of large data
  - Create your own application
  - Use specific libraries for data analysis in Agriculture
  - Connect to different data sources and move large datasets easily
  - Visualization features and custom frontends

- Easy usage of HPC + Cloud resources
  - Do not care about complex access mechanisms for storage and computation
  - Optimize HPC+Cloud combination for the user

- Default set of applications ready for the user
  - 6 re-usable scenarios
  - 3 re-usable pilots
  - Standardized APIs
Managing Large Data

- Try to minimize large data movement
  - Datasets sources evaluation model → Decide best source when available in more than one location
  - Use caching of datasets → Local repos for most used datasets
  - Try to keep computation close to the data → EUXDAT Cloud close to Mundi

- Optimize local storage resources
  - Be smart with caching
  - Try to predict datasets needs → Move earlier

- Example: existing tools for managing data

[Logos of RUCIO, MapServer, and MapProxy]
Data Analysis and Visualization Tools

http://www.euxdat.eu/
Overview of the Achievements (I)

- Requirements, Features and Pilots/Scenarios
  - Requirements collected from pilots and stakeholders
  - Two versions of the requirements, features and architecture
  - Definition of the pilots and their scenarios
  - First scenarios implemented and others ongoing
    - Open Land Use Map
    - Crop Monitoring
    - Agro-Climatic zones
Overview of the Achievements (II)

Technical Components

- **Cloud + HPC environments**
  - PaaS layer to host User’s platform
  - Elastic resources management
  - HPC+Cloud orchestration

- **User’s platform**
  - Data Connectors
  - Data Management
  - Development environment
  - Execution Orchestrator
  - Monitoring
  - Data Analytics tools (GRASS, Orfeo, etc...)
<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Data Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite Images</td>
<td>&gt;8PB</td>
</tr>
<tr>
<td>Sensors Data</td>
<td>4TB</td>
</tr>
<tr>
<td>Hyperspectral Images</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Meteo Data</td>
<td>337+1TB (9PBs)</td>
</tr>
<tr>
<td>Machine Monitoring</td>
<td>-</td>
</tr>
<tr>
<td>VGI Data</td>
<td>40GB</td>
</tr>
<tr>
<td>Other Vector Data</td>
<td>1GB (100GB)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Services</th>
<th>Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Connectors</td>
<td>6</td>
</tr>
<tr>
<td>Data Analytics Tools</td>
<td>5</td>
</tr>
</tbody>
</table>

http://www.euxdat.eu/
Challenges to be Addressed

- Data sources...
  - Analysis of Hyperspectral images from UAV
  - Data-in-motion analysis (compute at the edge) → Farm stations & machinery
- Plugins for Rucio (i.e. with the Data Catalogue, monitoring...)
- Parallelization of analysis for HPC
- Profiling and optimization of tasks execution + SLAs
- Marketplace setup and access control/billing mechanisms to implement a sustainable business model
- Involve more end users (hackatons, give access, etc...)

http://www.euxdat.eu/
Thank you for your attention

F. Javier Nieto
francisco.Nieto@atos.net

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 777549