CANDELA architecture for Data Mining and Data Fusion: a DIAS contribution

DLR
ATOS FR
ATOS ES
CANDELA main objective

CANDELA project main objective is to allow the **creation of value** from **Copernicus data** through the provisioning of **modelling analytics tools** given that the tasks of data collection, processing, storage and access will be provided by the **Copernicus Data and Information Access Service (DIAS)**.

Concept: **Data Science** to enable the successful **integration of heterogeneous datasets**, to support the definition and design of the **data transformation to information**, the use of taxonomies and elements of **ontology and semantics**, learning, KDD, annotation, data analytics.
Preliminaries

- In CANDELA, a special attention is given to **re-use and openness**.
  - building modules and frameworks on-top of available components
  - maximization of benefits from existing assets
  - making the solutions available to various user communities

- EOLib is an Image Information Mining system for TerraSAR-X GS:
  - processes, extracts, and accesses the content of EO products
  - generates higher-level abstractions and semantics
  - offers information mining services on the original corpus of EO products
  - provides KDD based on the EO content, metadata, semantic annotations

- In CANDELA we evolve selected modules from EOLib for COPERNICUS
The analytics modules

Source: CANDELA D2.1 Data Mining v1 v1.0
Data Mining and Fusion in CANDELA

Source: CANDELA D2.1 Data Mining v1 v1.0
Data Mining Architecture

Source: CANDELA D2.2 Data Mining v2 v1.0
Data Fusion Architecture

1. CANDELA Platform
   - Container: DMG-DF
     - Generate data model
   - Container: DBMS-Monetdb
     - Ingest data model to database

2. Internet
   - Jupyterlab
     - Quick-look images
   - REST-Service
     - Query: product metadata
       - Ingest: annotated semantics

3. Semantic annotation

Source: CANDELA D2.8 Data Fusion v2 v1.0
CreoDIAS: data connector

CreoDIAS provides access to metadata through a Data Finder API, to discover data:

- name
- platform
- instrument
- resolution
- productType
- cloudCover
- snowCover
- cultivatedCover
- ... etc.
CANDELA platform: architecture

- **Kubernetes**: an open-source container-orchestration system for automating application deployment, scaling, and management

- **Docker**: open-source OS-level virtualization to deliver software in packages called containers.

Source: CANDELA D3.5 System Integration and Validation Test Plan v1 v1.0
CANDELA platform: Jupyterlab

- **Launch_dmg_s1.jynb**
  - Visualize and choose Sentinel-1 product
  - Start data model generation to process Sentinel-1 data
- **Launch_dmg_s2.jynb**
  - Similar to ‘launch_dmg_s1’
- **Launch_datafusion.jynb**
  - to be tested
- **Monetdb.jynb**
  - Search and query in database
User frontend: Data Mining & Data Fusion

- User clicks buttons to operate the ‘cascade active learning’ system

<table>
<thead>
<tr>
<th>Query the database</th>
<th>Learn the training samples</th>
<th>Choose the grid level</th>
<th>Load the next batch of patches</th>
<th>Ingest into the database</th>
<th>Go to the previous image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go to the next image</td>
<td>Reset the current layer</td>
<td>Restart from the first layer</td>
<td>Show the positive and negative samples</td>
<td>Export annotations to .png files</td>
<td>Compute evaluation measures</td>
</tr>
</tbody>
</table>

- Quick-look panel:
  - one column for Data Mining
  - double column for Data Fusion

Source: CANDELA D2.8 Data Fusion v2 v1.0
User frontend: classification map

- **Image mining:**
  - Discover patterns in images
  - Semantic annotation based on a 3-level hierarchical semantic catalogue:
    - Agriculture
      - Cropland, greenhouses, pasture, plantations and vegetables, rice paddles, etc.
    - Industrial areas, etc.

- **Further data analysis:**
  - Semantic annotation distribution
  - Precision-recall evaluation

Source: CANDELA D2.1 Data Mining v1 v1.0
CANDELA D2.2 Data Mining v2 v1.0
CANDELA platform: MonetDB client

- Data mining in database
  - Metadata
    - mission
    - latitude
    - longitude
    - quick look image path
    - ...
  - Product features
    - Gabor linear moment (GLM)
    - color histogram (CHIS)
    - ...
  - Annotations
    - User discovered from images
Thank you for your attention

Wei Yao, Corneliu Octavian Dumitru, Mihai Datcu (DLR)
Jean-Francois Rolland, Fabien Castel (ATOS FR)
Jose Lorenzo (ATOS ES)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776193

www.candela-h2020.eu