Copernicus Access Platform Intermediate Layers Small Scale Demonstrator

# D5.6 Data Management Plan v3

| **Document Identification** | | | |
|---|---|---|---|
| **Status** | Final | **Due Date** | 31/10/2020 |
| **Version** | 2.0 | **Submission Date** | 19/02/2021 |

| | | | |
|---|---|---|---|
| **Related WP** | WP5 | **Document Reference** | D5.6 |
| **Related Deliverable(s)** | D1.1, D1.2, D5.7, D5.4, D5.5, D5.8 | **Dissemination Level (*)** | PU |
| **Lead Participant** | ATOS ES | **Lead Author** | Juan Alonso |
| **Contributors** | TerraNIS, SGIS, ATOS FR, TAS FR, IRIT | **Reviewers** | N/A |
| | | | N/A |


| **Keywords:** |
|---|
| Findable, Accessible, Interoperable and Reusable, DWH mechanism |


This document is issued within the frame and for the purpose of the CANDELA project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 776193. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

# Document Information

| List of Contributors | |
|---|---|
| Name | Partner |
| Juan Alonso, Jose Lorenzo | ATOS ES |
| Mohanad Albughdadi | TerraNIS |
| Anna Pulak-Siwiec, Agnieszka Bolibok | SGIS |
| Antonio Bartolini | TAS IT |
| Fabien Castel, A. HAUGOMMAR | ATOS FR |
| Michelle Aubrun | TAS FR |
| Ba Huy Tran | IRIT-CNRS |

| Document History | | | |
|---|---|---|---|
| Version | Date | Change editors | Changes |
| 0.1 | 01/10/2020 | J. Alonso (ATOS ES) | Draft version |
| 0.2 | 06/10/2020 | Mohanad Albughdadi (TerraNIS) | TerraNIS contributions, sections 3.1, 3.2 and 2.1 |
| 0.3 | 07/10/2020 | Anna Pulak-Siwiec (SGIS) | SGIS contributions, sections 3.3 and 3.4 |
| 0.4 | 08/10/2020 | Jean-Francois Rolland (ATOS FR) | Atos FR contributions, section 21 |
| 0.5 | 09/10/2020 | Juan Alonso (ATOS ES) | Consolidated version |
| 0.6 | 15/10/2020 | Juan Alonso (ATOS ES) | Updated version including comments from JF |
| 0.7 | 16/10/2020 | Juan Alonso (ATOS ES) | Updated version including comments from SGIS, TerraNIS and ATOS FR |
| 1.0 | 22/10/2020 | Jose Lorenzo (ATOS ES) | Final revision for submission |
| 1.1 | 18/02/2021 | Agnieszka Bolibok (SGIS), Mohanad Albughdadi (TerraNIS), Michelle Aubrun (TAS FR), Ba Huy Tran (IRIT), Juan Alonso (ATOS ES) | This version takes into account Review's comments (02/2021).<br>The following sections have been added: 3.1.6, 3.1.7, 3.2.6, 3.2.7, 3.3.6, 3.3.7, 3.4.6, 3.4.7 and 3.5. |
| 2.0 | 19/02/2021 | Jose Lorenzo (ATOS ES) | Final revision for submission |

| Quality Control | | |
|---|---|---|
| Role | Who (Partner short name) | Approval Date |
| Deliverable leader | Juan Alonso (ATOS ES) | 18/02/2021 |
| Quality manager | Juan Alonso (ATOS ES) | 18/02/2021 |
| Project Coordinator | Jose Lorenzo (ATOS ES) | 19/02/2021 |

# Table of Contents

# List of Figures

# List of Acronyms

| Abbreviation / acronym | Description |
|---|---|
| ADD | ADDitional datasets |
| D5.4 | Deliverable number 4 belonging to WP5 |
| DAP | Data Access Portfolio |
| DMP | Data Management Plan |
| DWH | Data WareHouse |
| EC | European Commission |
| ESA | European Space Agency |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| GRD | Ground Range Detected |
| IW | Interferometric Wide swath mode |
| ROI | Region of Interest |
| SAR | Synthetic Aperture Radar |
| SC | ScarSAR |
| SLC | Single Look Complex |
| ST | SpotLight |
| VHR | Very High-Resolution |
| WGS84 | World Geodetic System 1984 |
| WP | Work Package |

# Executive Summary

This deliverable corresponds with the third and final version of the Data Management Plan (DMP) and describes the updated DMP for the CANDELA project. The goal of this document is to provide an overview of all datasets used by the project.

The first DMP version  D5.4 [15]) followed already the guidelines of the template provided by the European Commission (Template Horizon 2020 Data Management Plan [1] [2]). The Data Management task (T5.3) is included within the Project Management and Coordination work package (WP5).

The second DMP (D5.5[16]) was submitted in October 2019 including updates of chapter "2.1 FAIR data", the CANDELA Project datasets (section 3) and information related to Section 4.

This third version includes the following changes:

The Data Management Plan chapter (2) has been updated to include details about the solution proposed by the project to publish new data (see section 2.1.1 Making data findable, including provisions for metadata). Section 2 also includes additional information to explain how to make data openly accessible (section 2.1.2) and interoperable (section 2.1.3).

Once the final validation for each use case took place, partners in charge of use cases (TerraNIS and SGIS) have updated the project datasets in section 3 CANDELA project datasets. The updated version submitted in February 2021 also includes "Output and Auxiliary Data" for each use case. In addition, section 3.5 has been added including information about the Nature 2000 use case.

Regarding Additional datasets, sections 4.3 and 4.4 have been added just to explain why it was not needed request of ADD in 2020.

# 1 Introduction

## 1.1 Purpose of the document

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the project [6]. Following the DMP template, this document provides for each dataset:

- Identifier and description
- Reference to existing suitable standards
- Description about data sharing and preservation

During the project lifetime, the document needs to be updated, so it has been planned the release of three versions of the CANDELA DMP.

- D5.4 DMP v1: the first version of the document based on the DMP template provided by the European Commission [1][2].
- D5.5 DMP v2: the DMP will evolve and will be updated in October 2019 (M18).
- D5.6 DMP v3: the final version of the document is planned in October 2020 (M30)

## 1.2 Relation to other project work

CANDELA project has planned reports about the request of Additional Datasets thorough the Data Warehouse mechanism managed by ESA. The different versions of the DMP will include the reference to these reports.

## 1.3 Structure of the document

This document is structured in four major chapters described below:

- **Chapter 2** presents the Data Management Plan (FAIR data principle)
- **Chapter 3** presents the CANDELA project Datasets
- **Chapter 4** presents Additional Datasets
- **Chapter 5** presents the conclusions

# 2 Data management Plan

This section has been configured following the DMP template [1][2] in order to achieve the FAIR principle for the project datasets.

## 2.1 FAIR data

According to the DMP template, the CANDELA research data should be “FAIR”: Findable, Accessible, Interoperable and Re-usable. This section describes the procedures to follow the FAIR principle.

### 2.1.1 Making data findable, including provisions for metadata

The CANDELA platform is built upon the CreoDIAS basis platform. Most of the dataset used in the context of CANDELA are or will be directly retrieved through API provided by CreoDIAS: Sentinel 1 and 2 images, Landsat images, other Copernicus data...

CreoDIAS provides a catalogue service containing information about:

- scenes/tiles
- orbits
- all available meta-data for individual scene and orbit
- link to quick-look images
- support for various processing levels (L1C, L2A, etc.)
- information about planned future acquisitions with a rich number of criteria (including at least satellite unit, instrument mode, polarization, geographical area, time window)

The catalogue service supports querying based on number of criteria:

- geographical area
  - bounding box (rectangle)
  - geometry (e.g. for agriculture fields)
- mission/sensor
- cloud coverage
  - based on scene meta-data
  - if more detailed cloud data are available, also based on location
- time interval
  - absolute based on "date" or "date and time" ranges defined using ISO8601 standard (from, to, from/to)
  - relative time intervals ("last week", "last month", last 2 months", "from last 6 months up to last 3 months", etc.)
  - advanced time intervals ("months May-August in from 1985 to 2016)
- mosaicking priority (most recent/least recent/least cloudy)

INSPIRE and GEOSS compatibility is ensured for relevant data-sets.

Dataset produced in the frame of CANDELA can also be catalogued to make them findable. The main solution proposed by CANDELA to publish new data is the semantic database included in the toolset provided by the IRIT. Results of the different processes (change detection, classification) proposed in CANDELA can be ingested in a PostGIS semantic database using the triplification process. This process transforms the data generated by other algorithms such as change detection and classification into semantic data. This data can be accessed in several ways: using the graphical user interface developed by IRIT, using the triplestore endpoint, or by semantic requests directly from the jupyter environment.

The classification data created using DLR tools is also stored in a specific database. This database can be accessed using the graphical interface provided by DLR or directly by executing SQL requests from the jupyter-lab.

Datasets produced in the framework of CANDELA and processed by the semantic search tool allow finding the corresponding satellite images used for the use case. A GeoServer [7] tool instance is used to run processing task on the platform and can be used to manage data.

The previous information will help making the data discoverable.

### 2.1.2 Making data openly accessible

Once the datasets are on the CANDELA platform, access permissions (who can access what) can be defined. Depending on the willingness of the data owner, a data sharing strategy will be drawn. Possible access permissions can be:

- Data without any restriction: all users can visualize, edit and download data without constraints.
- Protected data: rights will be defined for users in order to visualize, edit or download the data.
- Private data: datasets are not shared and only the user who produced them can access them.

The data stored in CANDELA databases (Strabon, and montDB) are accessible for all users. The data created by users inside their workspace remains private. A public shared space has been created, in this space any user can share its data or code.

### 2.1.3 Making data interoperable

Datasets on the CANDELA platform will be interoperable due to their compliance with international standards such as Web Feature Service (WFS), Web Map Service (WMS) and Web Coverage Service (WCS). GeoServer and OWSLib [9] Python library are used as implementation of these protocols.

All satellite any images will be in JPEG2000 or GeoTIFF format and they can be associated with metadata and provided through WMS services using the dedicated WMS lib library.

### 2.1.4 Increase data re-use (through clarifying licenses)

For open source data such as Sentinel images, used licenses are indicated in the metadata files. Users need to check these licenses before using or sharing the data. On the other hand, use strategies will be defined for data and services that will be produced as part of the project.

# 3 CANDELA project datasets

## 3.1 SUC1_Urbanization

The final validation of this sub-use case was conducted on the region of Bordeaux. The aim of this sub-use case is to study the changes related to urban expansion and agriculture. The region around Bordeaux is known for the continuous build-up of urban areas and richness of agricultural zones. The following datasets were used:

- Sentinel-2 images were used for the change detection pipeline developed by TAS FR.
- Sentinel-1 images were used for the change detection pipeline developed by TAS IT.
- Sentinel-1 and Sentinel-2 images were used for the data mining and data fusion modules developed by DLR.
- Sentinel-2 images are used for the semantic search tool developed by IRIT.

Please note that for the final validation, no VHR datasets are required from the DWH. Additionally, no in-situ data were used for this use case. The final output datasets are also described as well as the used auxiliary data sources.

### 3.1.1 Dataset reference and name

| Dataset | Summary |
|---|---|
| Sentinel-2 | Sentinel-2 satellite images are optical datasets produced as part of the EU Copernicus program.<br>These data are collected from Sentinel-2 mission that comprises a constellation of two polar-orbiting satellites.<br>Images coming from Sentinel-2 mission are characterized with a rich spectral resolution (13 bands), a high spatial resolution (10m panchromatic band) and a high acquisition frequency (5 days in cloud-free conditions). Hence, they are very suitable for monitoring variability in land surface.<br>The datasets are in ESA SAFE format and contain inside one tile with all spectral bands. |
| Sentinel-1 | Sentinel-1 satellite images are SAR datasets produced as part of the EU Copernicus program. The Sentinel-1 mission consists of two satellites operating day and night to perform C-band synthetic aperture radar imaging and allow an acquisition frequency of 6 days for the same place. Data are acquired according to different modes and the one that will be used for this use case is the Interferometric Wideswath mode (IW). This mode exists at a different level of correction, namely, the Single Look Complex (SLC) and the Ground Range Detected (GRD). |

### 3.1.2 Dataset description

| ROI | Dataset | Description |
|---|---|---|
| *Bordeaux* | *Sentinel-2* | • *Period: Images acquired in the period between 2016 and 2019.*<br>• *Surface: 500 km2.*<br>• *Coordinates: xMin,yMin -0.703984,44.7483 : xMax,yMax -0.46762,44.9188 (WGS84).*<br>• *S2 tilles: 30TXQ.* |
| | *Sentinel-1* | • *Period: Images acquired in the period between 2016 and 2019.*<br>• *Surface: 500 km2.*<br>• *Coordinates: xMin,yMin -0.703984,44.7483 : xMax,yMax -0.46762,44.9188 (WGS84).* |

### 3.1.3 Standards and metadata

| Dataset | Standards | Metadata |
|---|---|---|
| Sentinel-2<br>https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats | 13 JPEG2000 images, one for each band. | XML file |
| Sentinel-1<br>https://earth.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-product-formatting | Image data are stored in a Geotiff format. | XML file and Geotiff SAR image |

### 3.1.4 Data Sharing

Sentinel-1 and Sentinel-2 data are free of charge and can be downloaded from Copernicus Open Access Hub upon subscription and the CreoDIAS API.

### 3.1.5 Archiving and preservation

All the data described above will be conserved on the data server of the platform during all the project life time for demonstration purposes.

### 3.1.6 Output data

| Dataset | Used tool | Summary | Metdadata |
|---|---|---|---|
| Optical change detection maps | Optical change detection | Optical change detection maps were generated using Sentinel-2 images of the T30TXQ tile. The first map was obtained by processing two images acquired on the 2nd August 2017 and 23rd | Geotiff image |

| | | | |
|---|---|---|---|
| | | February 2019. The second map corresponds was obtained by processing two images acquired on the 23rd February 2019 and 22nd August 2019. Finally, a third map was generated using two images acquired on the 2nd August 2017 and 22nd August 2019. | |
| SAR change detection maps | SAR change detection | SAR change detection maps were generated using Sentinel-1 images around Bordeaux in France with a bounding box of this area is (xMin,yMin - 0.703984,44.7483: xMax,yMax - 0.46762,44.9188). Two change detection maps were generated using the VV and VH polarizations, respectively, of two images acquired in 2017 and 2019. | Geotiff image |
| Patch-level classification | Data mining tool | Two patch-classification maps were generated using two Sentinel-2 images of the T30TXQ acquired on the 2nd August 2017 and 22nd August 2019. | Geotiff image and SQL records |
| Patch-level classification | Data fusion tool | A patch classification map was generated using Sentinel-1 and Sentinel-2 images acquired on the 29th August 2019 and 22nd August 2019, respectively. | Geotiff image and SQL records |
| Semantic search datasets | Semantic search tool | Two datasets were generated using the semantic search tool with the NDVI of two Sentinel-2 images of the T30TXQ tile acquired on the 2nd August 2017 and 22nd August 2019 and the commune boundaries dataset that corresponds to the extent of the T30TXQ tile. | Accessed via semantic search modules |

### 3.1.7 Auxiliary data

| Dataset | Source | Summary | Metadata |
|---|---|---|---|
| Land cover classification | CESBIO<br>http://osr-cesbio.ups-tlse.fr/~oso/posts/2018-04-09-carte-s2-2017/<br>http://osr-cesbio.ups-tlse.fr/~oso/posts/2019-03-25-carte-s2-2018%20/ | An annual land cover classification map generated by the CESBIO laboratory in France and contains multi-class pixel classification. The maps of 2017 and 2018 were used | Geotiff |

| | | | |
|---|---|---|---|
| | | to validate this use case. | |
| Commune contours | https://www.data.gouv.fr/fr/datasets/decoupage-administratif-communal-francais-issu-d-openstreetmap/ | A vector file containing the boundaries of French communes. | Shapefiles |

## 3.2 SUC2_Vineyard

This sub-use case aims at assessing the damage level in vineyards caused by natural hazards such as frost and hail. For this purpose, the following datasets were identified for this sub-use case:

- Since the assessment of vineyards damage needs to be done as soon as the event happens, Sentinel 2 images will be used for this use case.

Please note the deviation of this section when compared to the first version of this document. The use of VHR satellite images will be challenging as it requires programming the acquisition before and after a natural disaster, which is difficult if not impossible. Additionally, no in-situ data were used for this use case. The final output datasets are also described as well as the used auxiliary data sources.

### 3.2.1 Dataset reference and name

| Dataset | Summary |
|---|---|
| Sentinel-2 | Sentinel-2 satellite images are optical datasets produced as part of the EU Copernicus program.<br>These data are collected from Sentinel-2 mission that comprises a constellation of two polar-orbiting satellites.<br>Images coming from Sentinel-2 mission are characterized with a rich spectral resolution (13 bands), a high spatial resolution (10m panchromatic band) and a high acquisition frequency (5 days in cloud-free conditions). Hence, they are very suitable for monitoring variability in land surface.<br>The datasets are in ESA SAFE format and contain inside one tile with all spectral bands. |

### 3.2.2 Dataset description

| ROI | Dataset | Description |
|---|---|---|
| *Bordeaux vineyards* | *Sentinel-2* | • *Period: two Images, the first one was acquired on 19th April 2017 and the second image on the 29th April 2017 (after and before a frost event).*<br>• *Surface: 1233 km².*<br>• *Coordinates: xMin,yMin -0.374589,44.5206 : xMax,yMax 0.0949954,44.7976 (WGS84).*<br>• *S2 tilles: 30TYQ.* |

### 3.2.3 Standards and metadata

| Dataset | Standards | Metadata |
|---|---|---|
| Sentinel-2<br>https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats | 13 JPEG2000 images, one for each band. | XML file |

### 3.2.4 Data Sharing

Sentinel-2 data are free of charge and can be downloaded from Copernicus Open Access Hub upon subscription and the CreoDIAS API.

### 3.2.5 Archiving and preservation

All the data described above will be conserved on the data server of the platform during all the project life time for demonstration purposes.

### 3.2.6 Output data

| Dataset | Used tool | Summary | Metadata |
|---|---|---|---|
| Optical change detection maps | Optical change detection | Optical change detection map was generated using Sentinel-2 images of the T30TYQ tile. This map was obtained by processing two images acquired on the 19th April and 29th April 2019. | Geotiff images |
| SAR change detection maps | SAR change detection | SAR change detection maps were generated using Sentinel-1 images of the bounding box of the ROI as described in Section 3.11. Two change detection maps were generated using the VV and VH polarizations, respectively, of two images acquired in 19th April and 29th April 2017. | Geotiff images |
| Patch-level classification | Data mining tool | A patch-classification map was generated using a Sentinel-2 image of the T30TYQ acquired on the 19th April. | Geotiff images and SQL records |
| Semantic search datasets | Semantic search tool | Two datasets were generated using the semantic search tool with the NDVI of two Sentinel-2 images of the T30TYQ tile acquired on the 19th and 29th April 2017 and the parcel boundaries dataset that corresponds to the extent of the ROI. | Accessed via the semantic search modules |

### 3.2.7 Auxiliary data

| Dataset | Source | Summary | Metadata |
|---|---|---|---|
| Land cover classification | CESBIO http://osr-cesbio.ups-tlse.fr/~oso/posts/2018-04-09-carte-s2-2017/ | An annual land cover classification map generated by the CESBIO laboratory in France and contains multi-class pixel | Geotiff |

| | | | |
|---|---|---|---|
| | | classification. The map of 2017 was used to validate this use case. | |
| Parcel contours | https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-culturaux-et-leur-groupe-de-cultures-majoritaire/ | A vector file containing the boundaries of the vineyard parcels. | Shapefiles |

## 3.3 SUC3_ Forest-Disaster

This case concerns occurring disasters in forests (windfalls) on the example of RDLP Toruń in 2017. The study plans to analyze the extent of the occurrence of the disaster, study the area and extent of damage. An analysis of the activities carried out by the forestry services in the areas affected by the disaster is also carried out.

Satellite images acquired using different sensors were used for this sub-use case.

- Sentinel-2 and Sentinel 1 datasets were used to detect information about areas that were affected by windthrow.
- In the final tools validation, no VHR data were used.
- No in-situ data were used for this use case. Validation was carried out on the results of analyzes based on VHR data.

### 3.3.1 Dataset reference and name

| Dataset | Summary |
|---|---|
| Sentinel-1 | First satellite in Copernicus Program with C-SAR instrument which can operate in four imaging modes with different resolution (down to 25m). The mission is composed of two satellites, sharing the same orbital plane. |
| Setninel-2 | Sentinel-2 satellite images are optical datasets gathered within the EU Copernicus program.<br>The Sentinel-2 mission from which the data originates consists of a constellation of two satellites orbiting the Earth. Characteristics of Sentinel-2 based images are: rich spectral resolution (13 bands), a high spatial resolution (10m panchromatic band) and a high acquisition frequency (5 days in cloud-free conditions). |

### 3.3.2 Dataset description

| ROI | Dataset | Description |
|---|---|---|
| *Torun_1 and Torun_2* | *Sentinel-1* | • *Period: Images acquired just before 12.08.2017 and after*<br>• *Surface: 860 km2.*<br>• *Coordinates: xMin,yMin 17.2997, 53.651802 : xMax,yMax 17.810054, 53.888559 (WGS84)*<br>• *S2 tile: 33UXV and 34UCE* |
| | *Sentinel-2* | • *Period: Images acquired just before 12.08.2017 and after*<br>• *Surface: 860 km2.*<br>• *Coordinates: xMin,yMin 17.2997, 53.651802 : xMax,yMax 17.810054, 53.888559 (WGS84)* |

### 3.3.3 Standards and metadata

| Dataset | Standards | Metadata |
|---|---|---|
| Sentinel-1<br>https://earth.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-product-formatting | Image data are stored in a Geotiff format. | XML file, geotiff SAR image & GTiff image |
| Sentinel-2<br>https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats | 13 JPEG2000 images, one for each band. | XML file |

### 3.3.4 Data Sharing

Data collected from Copernicus program: Sentinel-1 and Sentinel-2 can be download from Copernicus Open Access Hub.

### 3.3.5 Archiving and preservation

The datasets mentioned above will be stored on the platforms data server while project life cycle. However, due to the commercial license to use certain data, access to certain data sets will be limited.

### 3.3.6 Output data

| Dataset | Used tool | Summary | Metadata |
|---|---|---|---|
| Optical change detection maps | Optical change detection | Optical change detection map was generated using Sentinel-2 images of the T33UVX tiles. This map was obtained by processing images acquired on the 01st May and 28th September 2017. | Geotiff images |
| SAR change detection maps | SAR change detection | SAR change detection maps were generated using Sentinel-1 images of the bounding box of the ROI as described in Section 3.3.2. Change detection maps were generated using the VV and VH polarizations, respectively, of images acquired 5th August and 17th August 2017 | Geotiff images |
| Patch-level classification | Data mining tool | A patch-classification map was generated using a Sentinel-2 image of the T33UVX and T34UCE the 28th September 2017. | Geotiff images and SQL records |

| | | | |
|---|---|---|---|
| Semantic search datasets | Semantic search tool | Two datasets were generated using the semantic search tool with the NDVI of two Sentinel-2 images of the T33UVX and T34UCE tiles acquired on the 01st April and 16th October 2017 and the parcel boundaries dataset that corresponds to the extent of the ROI. | Accessed via the semantic search modules |

### 3.3.7 Auxiliary data

| Dataset | Source | Summary | Metadata |
|---|---|---|---|
| Aerial images | SmallGIS | Aerial Data acquired for windthrown damage assessment in 2017. Data were acquired and orthorectified and analyzed by SmallGIS to produce maps Representing the percentage of damage in a given area, GSD 10 cm. | Geotiff |
| BDL - Forest Data Bank | https://www.bdl.lasy.gov.pl/portal/ | A vector file containing the boundaries an attributes of specified forest areas the information used concerned the density and age of the trees | Shapefiles |
| Orthophoto (Geoportal) | Head Office of Geodesy and Cartography, www.geoportal.gov.pl | Additional photointerpretation for verification purposes (orthophotomaps) | WMS, WMTS, Shapefiles |
| Administrative regions and cadastral parcels | Head Office of Geodesy and Cartography, www.geoportal.gov.pl | Administrative regions and cadastral parcels | Shapefiles |

## 3.4 SUC4_ Forest-Health

This case concerns tree stand analyses during the bark beetle invasion in Białowieża. The analyses will include both large-scale and low-scale studies of forest condition.

The following datasets were identified and used for this sub-use case:

- Sentinel-2 for Optical change detection, Data mining, Data Fusion and Semantic Search,
- Sentinel-1 for SAR change detection.

Independent data for validation was obtained for this scenario from ForBioSensing Project and for that reason VHR data was not used and acquired. No direct in-situ data were used for this use case.

### 3.4.1 Dataset reference and name

| Dataset | Summary |
|---|---|
| Sentinel-1 | First satellite in Copernicus Program with C-SAR instrument which can operate in four imaging modes with different resolution (down to 25 m). The mission is composed of two satellites, sharing the same orbital plane. |
| Setninel-2 | Sentinel-2 satellite images are optical datasets gathered within the EU Copernicus program.<br>The Sentinel-2 mission from which the data originates consists of a constellation of two satellites orbiting the Earth. Characteristics of Sentinel-2 based images are: rich spectral resolution (13 bands), a high spatial resolution (10m panchromatic band) and a high acquisition frequency (5 days in cloud-free conditions). |

### 3.4.2 Dataset description

| ROI | Dataset | Description |
|---|---|---|
| *Białowieża* | *Sentinel-1* | • *Period: Images acquired from 2015 to 2018 in summer months*<br>• *Surface: 385 km2.*<br>• *Coordinates: xMin,yMin 23.68852, 52.674583 : xMax,yMax 23.949071, 52.870889 (WGS84)*<br>*S2 tile: 34UFD* |
| | *Sentinel-2* | • *Period: Images acquired from 2015 to 2018 in summer months*<br>• *Surface: 385 km2.*<br>*Coordinates: xMin,yMin 23.68852, 52.674583 : xMax,yMax* |
| *Niepolomice* | *Sentinel-1* | • *Period: Images acquired in 2020 in summer months*<br>• *Surface: 335 km2.*<br>• *Coordinates: xMin,yMin 20.1739, 49.9479 : xMax,yMax 20.4784, 50.0880 (WGS84)*<br>• *S2 tile: 34UDA* |
| | *Sentin* | • *Period: Images acquired in 2020 in summer months*<br>• *Surface: 335 km2.* |

| | | |
|---|---|---|
| | *el-2* | • *Coordinates: xMin,yMin 20.1739, 49.9479 : xMax,yMax 20.4784, 50.0880 (WGS84)*<br>• *S2 tile: 34UDA* |

### 3.4.3 Standards and metadata

| Dataset | Standards | Metadata |
|---|---|---|
| Sentinel-1<br>https://earth.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-product-formatting | Image data are stored in a Geotiff format. | XML file, geotiff SAR image & GTiff image |
| Sentinel-2<br>https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats | 13 JPEG2000 images, one for each band. | XML file |

### 3.4.4 Data Sharing

Data collected from Copernicus program: Sentinel-1 and Sentinel-2 can be download from Copernicus Open Access Hub.

### 3.4.5 Archiving and preservation

The datasets mentioned above will be stored on the platforms data server while project life cycle.

### 3.4.6 Output data

| Dataset | Used tool | Summary | Metadata |
|---|---|---|---|
| Optical change detection maps | Optical change detection | Optical change detection map was generated using Sentinel-2 images of the T34UFD and T34UDA tiles. For T34UFD the maps were obtained by processing images acquired in the years 2015,2016,2017 and for T34UDA in 2020. | Geotiff images |
| SAR change detection maps | SAR change detection | SAR change detection maps were generated using Sentinel-1 images of the bounding box of the ROI as described in Section 3.4.2. For T34UFD change detection maps were generated using the VV and VH polarizations, respectively, acquired in the years 2015,2016,2017, and for T34UDA in 2020. | Geotiff images |
| Patch-level | Data mining | A patch-classification maps were | Geotiff images and SQL |

| | | | |
|---|---|---|---|
| classification | tool | generated using a Sentinel-2 images of the T34UDA acquired in 2020 in Niepołomice area. | records |
| Semantic search datasets | Semantic search tool | The datasets were generated using the semantic search tool with the NDVI of Sentinel-2 images of the T34UDA acquired in 2020 in Niepołomice area. | Accessed via the semantic search modules |

### 3.4.7 Auxiliary data

| Dataset | Source | Summary | Metadata |
|---|---|---|---|
| ForBioSensing | Forest Research Institute<br>http://www.forbiosensing.pl/start | Additional information about tree crowns. | Shape |
| BDL - Forest Data Bank | https://www.bdl.lasy.gov.pl/portal/ | A vector file containing the boundaries an attributes of specified forest areas<br>the information used concerned the density and age of the trees | Shapefiles |
| Orthophoto (Geoportal) | Head Office of Geodesy and Cartography, www.geoportal.gov.pl | Additional photointerpretation for verification purposes (orthophotomaps) | WMS, WMTS, |
| Administrative regions and cadastral parcels | Head Office of Geodesy and Cartography, www.geoportal.gov.pl | Administrative regions and cadastral parcels | Shapefiles |

## 3.5 SUC5_Natura 2000

This use case consists in monitoring Natura 2000 areas in France, which are areas designated to conserve habitats and protect species. This use case has been added to demonstrate the capacity of CANDELA tools, above all the optical change detection tool, to deal with a big amount of data. Indeed, Natura 2000 has 1567 sites throughout the entire French land territory, that is mean an area of 98 Sentinel-2 tiles to monitor.

### 3.5.1 Dataset reference and name

| Dataset | Summary |
|---|---|
| Sentinel-2 | Sentinel-2 satellite images are optical datasets produced as part of the EU Copernicus program.<br>These data are collected from Sentinel-2 mission that comprises a constellation of two polar-orbiting satellites.<br>Images coming from Sentinel-2 mission are characterized with a rich spectral resolution (13 bands), a high spatial resolution (10m panchromatic band) and a high acquisition frequency (5 days in cloud-free conditions). Hence, they are very suitable for monitoring variability in land surface.<br>The datasets are in ESA SAFE format and contain inside one tile with all spectral bands. |

### 3.5.2 Dataset description

| ROI | Dataset | Description |
|---|---|---|
| *France* | *Sentinel-2* | • *Period: between April and May 2019*<br>• *Coordinates: Polygon([(2.5, 51.1), (8.3, 49.0), (8.0, 43.7), (3.3, 42.2), (-2.0, 43.3), (-5.0, 48.4)]) in WGS84*<br>• *Surface: 98 pairs of S2 tiles → 2 x 98 x 110x110km$^2$* |

### 3.5.3 Standards and metadata

| Dataset | Standards | Metadata |
|---|---|---|
| Sentinel-2<br>https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats | 13 JPEG2000 images, one for each band. | XML file |

### 3.5.4 Data Sharing

Sentinel-2 data are free of charge and can be downloaded from Copernicus Open Access Hub upon subscription and CreoDIAS API in our case.

### 3.5.5 Archiving and preservation

All the data described above will be conserved on the data server of the platform during all the project life time for demonstration purposes.

### 3.5.6 Output data

| Dataset | Used tool | Summary | Metdadata |
|---|---|---|---|
| Optical change detection maps | Optical change detection | Optical change detection maps were generated using pairs of Sentinel-2 images that represent exactly the same field of view at two different times.<br>Optical change detection maps are geoTiff rasters with the same size and geo-reference as the images on which they have been calculated, and whose pixel values range between 0 and 1 in order to represent the probability that a change has occurred between the acquisition dates of the two images. | GeoTiff image |
| RBF database | Triplification tool | RBF database was generated using the change detection maps described above and the Natura 2000 database described below.<br>RBF database contains change detection labels (Very Low_ChangeDetect, Low_ChangeDetect, Middle_ChangeDetect and High_ChangeDetect) for each Natura 2000 site, as well as the metadata (coordinates of Natura 2000 site, ID of Sentinel-2 data used, acquisition date, …) | RBF database |

### 3.5.7 Auxiliary data

| Dataset | Source | Summary | Metadata |
|---|---|---|---|
| Natura 2000 database | https://www.eea.europa.eu/data-and-maps/data/natura-2000-eunis-database | This database is a compilation of data on ecological sites provided by member states to the European Commission. The database contains two categories of data:<br>• Descriptive data: The data are based on the information that national authorities have submitted for each Natura 2000 site<br>• Spatial data: The data provides the code and the name of the sites and the coordinates of the polygon or the multi-polygon sites | shapefile or json file |

# 4 Additional datasets

As part of the datasets used in the project, this section provides information about the use of Additional datasets (ADD) following the process stablished by ESA through the Data Warehouse mechanism (DWH). Detailed information about the project request and the use of ADD is available in deliverables D5.7 [5][10] and D5.8 [12] respectively.

## 4.1 Request of additional datasets 2019

CANDELA project requested Additional datasets (ADD) for 2019. The official request was included in the project deliverable D5.7 [5][10] (v1.0 submitted in September 2018 and v2.0 submitted in February 2019). Below is presented the final list of Additional datasets requested in February 2019.

| Project Name : CANDELA (776193) | | Project Coordinator: Jose Lorenzo (Atos Spain) | | |
|---|---|---|---|---|
| Category: Projects | | | | 2019 |
| Core dataset | code | | | Total sqkm |
| Archive_standard_Optical_VHR1 | ADD_011a | | | 6.240 |
| Archive_standard_SAR_VHR1 | ADD_015a | | | 6.240 |
| New acquisition_standard_Optical_VHR1 | ADD_012a | | | 3.685 |
| New acquisition_standard_SAR_VHR1 | ADD_016a | | | 3.685 |

Figure 1: Consolidated Request February 2019 (D5.7)[10]

The project request was validated internally by the European Space Agency (ESA) in August 2019 when the Agency published version 2.6[11] of the Data Access Portfolio Document (DAP). As it is showed in Figure 2, the validation included datasets requested for both versions of D5.7 [5][10].

| Project | Dataset Title | Dataset ID | Quota (km2) |
|---|---|---|---|
| CANDELA | Archive_standard_Optical_VHR1 | D2_MG2b_CAND_011a | 700 |
| CANDELA | Archive_standard_Optical_VHR2 | D2_MG2b_CAND_011b | 700 |
| CANDELA | Archive_standard_SAR_VHR1 | D2_MG1_CAND_015a | 1,270 |
| CANDELA | A/rchive_standard_SAR_VHR2 | D2_MG1_CAND_015b | 17,500 |
| CANDELA | New acquisition_standard_Optical_VHR1 | D2_MG2b_CAND_012a | 600 |
| CANDELA | New acquisition_standard_Optical_VHR2 | D2_MG2b_CAND_012b | 600 |
| CANDELA | New acquisition_standard_SAR_VHR1 | D2_MG1_CAND_016a | 885 |
| CANDELA | New acquisition_standard_SAR_VHR2 | D2_MG1_CAND_016b | 15,000 |

Figure 2: Assigned quota DAP version August 2019[11]

## 4.2 DWH use for 2019

The use of Additional datasets was reported in D5.8[12] following the two use cases proposed in CANDELA:

- Use case 1: Agriculture and Macro-economics (see deliverables [3][13])
- Use case 2: Forest Health Monitoring (see deliverables [4][14])

D5.8[12] describes the use of ADD in the project focusing in the use for validations purposes. Some examples include the use of VHR images obtained from the DWH mechanism to validate using photo interpretation. In addition, VHR images could also be used to verify the results obtained by machine

learning algorithms developed in the project. Sections 3 and 4 of the aforementioned report D5.8[12] include detailed information of the use of additional datasets in the project.

## 4.3 Request of additional datasets 2020

As it was reported in D5.8[12] (DWH use for 2019), the use of ADD datasets in CANDELA was mainly focused in the validation process of the two project use cases.

Taking into account that the developed analytic tools in the project are dedicated to Copernicus data, it was decided that there is no need to order ADD datasets from DWH mechanism for 2020. The same explanation was included in the formal document D5.9[17].

## 4.4 DWH use for 2020

As it was informed in D5.9[17] (see also section 4.3 Request of additional datasets 2020), there was not needed to order ADD datasets from DWH mechanism for 2020. Therefore, D5.10 [18], which was devoted to report on the usage of DWH for 2020, it was submitted formally as an empty document.

# 5 Conclusions

The third and final version of the DMP (this document) includes updated information of the CANDELA DMP v2 released in October 2019[16]. Main changes in this version are listed below:

- Section 2 includes now additional details to explain how the dataset can also be catalogued to make them findable. More details have been added to clarify how to make data openly accessible and interoperable.
- The main information included in section 3 has been updated. Project datasets of the use cases have been updated after the final validation of each sub-use case. Nature 2000 use case has been also added in section 3.5.
- Although it was not needed to order ADD datasets for 2020, section 4 has also been updated giving the appropriate justification.

Following the same structure of the previous versions, the document describes the bold lines in the data management plan of CANDELA project, including how to make the project data findable, accessible, interoperable and reusable. Still, this document deals mainly with input data, however the updated version submitted in February 2021 also includes “Output data and Auxiliary data” (sections 3.x.6 and 3.x.7 for each use case). As a matter of fact, CANDELA does not aim at producing data, but at delivering services, and therefore, in general, there is no need for long term data preservation.

The use cases proposed in this project made use of the available Sentinel datasets that are characterized by a fine spatial and temporal resolutions. Additionally, these datasets are available free of charge, which guarantees their accessibility and the ability to reproduce the results obtained in the project. Moreover, the interoperability of the produced datasets (from change detection and data mining/data fusion) follow standard formats, which allows them to be easily integrated to WMS services for example.

The data produced on the CANDELA platform are also made available in a database. The triplification process ensures that the information produced by change detection algorithms or the classification tool is stored in a semantic database. Thus, the data in accessible in the database and can be requested through semantic requests. IRIT has also developed a graphical interface to explore this database. It allows to access to the data with low technical skills as the user does not have to write requests to the database.

# References


[1] DMP template, Annex 1; https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template

[2] Template Horizon 2020 DATA MANAGEMENT PLAN (DMP) http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx

[3] M. Albughdadi (2018), D1.1 Use Case #1 Requirements v1 v1.0, Deliverable of the CANDELA project

[4] A. Pulak-Siwiec (2018), D1.2 Use Case #2 Requirements v1 v1.0, Deliverable of the CANDELA project

[5] J. Lorenzo (2018), D5.7 CANDELA DWH request for 2019 (v1), Deliverable of the CANDELA project

[6] CANDELA Annex 1 to the Grant Agreement (Description of the Action) - NUMBER 776193 - Part A & Part B

[7] GeoServer data management, https://docs.geoserver.org/stable/en/user/data/webadmin/layers.html, retrieved 17/12/2018

[8] GeoNetwork open source catalogue, https://geonetwork-opensource.org/, retrieved 17/12/2018

[9] OWSLib Python library for OGC standards, https://geopython.github.io/OWSLib/, retrieved 17/12/2018

[10] J. Lorenzo (2019), D5.7 CANDELA DWH request for 2019 (v2), Deliverable of the CANDELA project

[11] G. Ottavianelli. Copernicus Space Component Data Access Portfolio (DAP v2.6): Data Warehouse 2014 – 2020. 31/08/2019, Issue/Revision 2.6 [Online]. Available: https://spacedata.copernicus.eu/documents/12833/14545/DAP+Document+-+current/c2449218-3ed9-434a-b32c-edfbb95b9362

[12] J. Alonso (2019). D5.8 DWH use for 2019 v1.0, Deliverable of the CANDELA project.

[13] M. Albughdadi (2019), D1.5 Use Case #1 Requirements v2 v1.0, Deliverable of the CANDELA project

[14] A. Pulak-Siwiec (2019), D1.6 Use Case #2 Requirements v2 v1.0, Deliverable of the CANDELA project

[15] J. Alonso (2018), D5.4 Data Management Plan v1, Deliverable of the CANDELA project

[16] J. Alonso (2019), D5.5 Data Management Plan v2, Deliverable of the CANDELA project

[17] 2019, D5.9 DWH request for 2020, Deliverable of the CANDELA project

[18] 2020, D5.10 DWH use for 2020, Deliverable of the CANDELA project