Ref. Ares(2020)5984118 - 27/10/2020

Copernicus Access Platform Intermediate Layers Small Scale Demonstrator

D1.11 Big data shift impact assessment report

Document Identification			
Status	Final	**Due Date**	31/10/2020
Version	1.0	**Submission Date**	27/10/2020

Related WP	WP1	**Document Reference**	D1.11
Related Deliverable(s)	All in general; D1.5; D1.6; D1.7; D1.8	**Dissemination Level (*)**	PU
Lead Participant	TAS France	**Lead Author**	Marie Desplas
Contributors	Romain Hugues (TAS FR), Michelle Aubrun (TAS FR), Marie Desplas (TAS FR), Mohanad Albughdadi (TerraNIS), Agnieszka Bolibok (SGIS)		
Reviewers	Jose Lorenzo Mon (ATOS ES), Juan Andres Alonso (ATOS ES)		

Keywords:
Big data technologies, cloud infrastructure, user capacity to exploit the Copernicus Data, create innovative applications, generate economic value, fulfilment of the project initial objectives

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.

(*) Dissemination level: **PU**: Public, fully open, e.g. web; **CO:** Confidential, restricted under conditions set out in Model Grant Agreement; **CI:** Classified, **Int** = Internal Working Document, information as referred to in Commission Decision 2001/844/EC.

Document Information

List of Contributors	
Name	Partner
Marie Desplas	Thales Alenia Space in France
Michelle Aubrun	Thales Alenia Space in France
Romain Hugues	Thales Alenia Space in France
Mohanad Albughdadi	TerraNIS
Agnieszka Bolibok	SGIS

Document History			
Version	Date	Change editors	Changes
0.0	07/09/2020	Marie Desplas (TAS FR)	TOC of the document
0.1	09/09/2020	Mohanad Albughdadi (TerraNIS)	Contribution to chapters 2, 3 and 4 for agriculture
0.2	09/10/2020	Agnieszka Bolibok (SGIS)	Contribution to chapters 2, 3 and 4 for forestry
0.3	16/10/2020	Michelle Aubrun (TAS FR)	Contribution to chapters 2 for Natura 2000 scenario presentation
0.4	19/10/2020	Marie Desplas (TAS FR)	First draft of the document
0.5	20/10/2020	Mohanad Albughdadi (TerraNIS)	Review the document
0.6	21/10/2020	Michelle Aubrun (TAS FR)	Integrate last modifications
0.7	26/10/2020	Juan Alonso (ATOS ES)	Quality assessment
1.0	27/10/2020	Jose Lorenzo (ATOS ES)	Final revision before submission

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Michelle Aubrun (TAS FR)	21/10/2020
Quality manager	Juan Alonso (ATOS ES)	26/10/2020
Project Coordinator	Jose Lorenzo (ATOS ES)	27/10/2020

Table of Contents

List of Figures

List of Acronyms

Abbreviation / acronym	Description
DIAS	Data and Information Access Services
EO	Earth Observation
GIS	Geographic Information System
GUI	Graphical User Interface
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared Radiation
PU	Public
RGB	Red-Green-Blue
SNAP	Sentinel Application Processing
UAV	Unmanned Aerial Vehicle
WP	Work Package

Executive Summary

The report explains CANDELA impact assessment on business thanks to big data shift.

The CANDELA project objective is to bridge the gap between big data technology and the Earth Observation data user community.

During the project, our consortium integrates already existing building blocks into a homogeneous, powerful and operational platform which enable users to quickly use, manipulate, explore and process Copernicus data.

Big data technologies associated to cloud infrastructure and a powerful platform have positive impact on the user capacity to exploit the Copernicus Data. This allows to improve societal and environmental impact and to generate economic value.

Big data also facilitates the analysis of large volume of Earth Observation Data to create innovative applications like for example to monitor the ecosystem health and evaluate the impact of the climate on the protected ecosystems or to monitor urban expansion, etc. Copernicus imagery combined to big data technologies has a strong potential of development in all the domains for which the earth observation is suitable: urban monitoring, coastal and marine exploitation and preservation, energy exploitation and implementation, environment monitoring or security operations.

All these new technologies and the range of applications that are made possible largely contribute to numerous economic, societal and environmental impacts. From the new applications presented in Tasks 1.1, 1.2 and the big data use case, this document summarizes the estimated potential impacts.

1 Introduction

The CANDELA work package 1 (WP1): "Big Data Shift for EO users" organizes the needs of end-users and the validation of these requirements on the CANDELA platform integrating already existing building blocks from the consortium partners.

This deliverable reports the impact assessment of the big data shift for the Earth Observation data user community and includes the project use cases and general impacts analysis among others on :

- The impact of the user capacity to exploit the Copernicus Data analysis including agricultural and forestry use cases.
- The possibility to compute change detection on large volume of Copernicus data associated to various data, based on a French use case to detect Natura 2000 zones evolution
- Emerging innovative applications enabled by new capacities with an analysis on agricultural and forestry new applications that can be developed
- An estimation of the economic value generated by these big data exploitation possibilities through agricultural and forestry CANDELA platform validations.

The materials presented in this document were elaborated with the active participation and contribution of all partners.

1.1 Purpose of the document

The purpose of the document is to explain the benefits and limitations of the CANDELA integrated platform for the Earth Observation data user community.

It presents the CANDELA end-users needs and today work compared to this same job done with the CANDELA platform. It also generalises this analysis to global impact assessment of the big data technologies for Earth Observation data users.

A new use case is presented to illustrate the capabilities of the platform and its impacts on a large volume of Copernicus data associated to various data. It is a Natura 2000 scenario run on the entire France country.

This document contains an analysis of what could be developed as innovative applications in the agricultural or forestry domains, as well as a general analysis innovation.

Finally, all this project experience is evaluated to analyse the economic value, societal and environmental impacts allowed by the big data shift of CANDELA project.

1.2 Relation to other project work

This deliverable evaluates the impact of the CANDELA project and the big data in general regarding the end-users needs and today work. This task in then related to all the WP1 tasks describing the end-users needs and their validation of the CANDELA platform for the day to day job.

1.3 Referenced documents

This deliverable is also based on the Copernicus Market report on Copernicus data exploitation and their impacts.

1.4 Structure of the document

The content of this deliverable is structured into three sections as follows:

- **Section 1: Introduction**
- **Section 2: User capability impact to exploit the Copernicus Data**
- **Section 3: Innovative applications creation**
- **Section 4: Economic value generation:** includes also a societal and environmental value analysis
- **Section 5: Conclusion**

2 User capability impact to exploit the Copernicus Data

2.1 User capacity impact to exploit the Copernicus Data on Agricultural domain

TerraNIS provides services in three main domains, namely, agriculture, viticulture and environment. These services rely on images coming from multiple sources. For instance, SPOT6/7 images are widely used for precision agriculture and viticulture. These images are characterized by a high spatial resolution (better than the spatial resolution of Sentinel-2 images). Hence, they can provide precise information. In the framework of TerraNIS' services, SPOT6/7 images are scheduled to be acquired at certain time instants of the growing season especially for viticulture. However, these images are not adapted for other services that require a high availability of satellite images such as irrigation and crop monitoring. Moreover, environmental applications require aggregating data from multiple years, which can be economically challenging using SPOT6/7. For these applications, Sentinel-2 images are more adapted as they have high frequency, cover larger geographical regions and are free of charge.

Multiple sources of Sentinel-2 data are used in TerraNIS such as the Scihub [2], the ONDA DIAS[1] and google earth engine, among others. Depending on the processing level (L1C or L2A), appropriate pre-processing chains are used. For L1C, two pre-processing chains are used. The first one uses the Overland software, which provides tools for atmospheric correction, cloud detection, etc. The second pre-processing chain is provided in the framework of the Sentinel Application Processing (SNAP), which provides the same capabilities as in the Overland software. For L2A image, pre-processing tools developed at TerraNIS are used. However, images from Google Earth Engine do not require external pre-processing chains as they are pre-processed and the user can also perform certain processing on these images such as the Normalized Difference Vegetation Index (NDVI), etc.

The use of Sentinel-1 data is quite limited at TerraNIS when compared to Sentinel-2 as they are more complex to use and interpret especially in the context of agriculture. Sentinel-1 data are more used for environmental services such as urban land cover detection and for research purposes. We rely on tools provided by external sources for pre-processing such as SNAP and PEPS [3].

Regarding the resources required to pre-process Sentinel data, the human factor is almost negligible as most of the available tools are automatized. However, a proper computation environment is needed especially when processing temporal sequences. It also worth noting that image retrieval is the most time-consuming part due to the inconsistency of the availability of Copernicus data from different sources.

For the time being, change detection techniques are not directly exploited in commercial offers proposed by TerraNIS. It is performed most of the time by performing a pre and post classification and then comparing their results. This process can lead to errors coming from the model performance and are not adapted for a precise change detection.

Large scale processing can be required for certain applications such as crop monitoring. For instance, change detection algorithms can be of interest for this kind of applications. We hope that with in-situ data, these changes can be correlated with agricultural anomalies such as water stress, insects, etc.

The CANDELA platform and the proposed analytics tools allow integrating new techniques to the existing offers of TerraNIS that bring an added-value and achieve good performance in the analysis. Additionally, these tools allow a large-scale processing, which requires a lot of resources for an SME

[1] https://www.onda-dias.eu

like TerraNIS. The use of these tools will allow us to concentrate the work on our agriculture and viticulture expertise and not on the infrastructure and IT parts. Nonetheless, these techniques need to be fine-tuned for each application and additional post-processing needs to be carried out to extract information adapted for each service.

One of the main limitations of the platform is the lack of graphical user interfaces (GUI). Most remote sensing technicians are used to work with software with a GUI such as SNAP, QGIS, ArcGIS, etc., where they can run processing steps without the need to interact with code. If the tools are to reach more user, GUIs will be mandatory.

2.2 User capacity impact to exploit the Copernicus Data on Forestry domain

The offer of SmallGIS includes, among others, dedicated remote sensing analyses as well as the development and GIS IT systems with integrated repositories of remote sensing data and analytical algorithms. Services offered by SmallGIS are mostly analyses or mechanisms dedicated to individual clients, developed on the basis of announced tenders. Depending on the client's expectations, analyses are performed on a defined time series of data or provided mechanisms that allow for the execution of an analysis on any set of data with defined parameters.

For searching and retrieving images from Copernicus Data Base we mostly use Copernicus Open Access Hub, Sentinel HUB, Earth explorer, dedicated QGIS extensions or Google Earth Engine. In the case of data that require preprocessing, we most often use generally available SNAP software.

Both in the case of data download and preprocessing, for small areas and data volumes, operations are performed manually in software that provides the appropriate GUI (like SNAP or QGIS). For larger volumes of data, scripts that automate individual steps, created by SGIS, are used. Depending on the number of data, the number of processing and the degree of their complexity scripts run locally on single computers or servers with higher computing power.

Change detection analysis are performed manually using tools available in software such as ArcGIS, QGIS, SNAP, ENVI (for small data volumes or testing and research purposes). In cases where not only the results of analyzes are delivered to the client, but also the tools to conduct them, SGIS solutions based on machine learning are increasingly used.

Copernicus data is most often used in environmental analyzes. Currently, for the needs of one of SGIS clients, the following analyzes based on Sentinel data are developed:

- determining the degree of humidity of the soil in different phases (periods) of vegetation
- initial determination of dynamics and range of invasive species succession
- determining the extent and dynamics of floods
- detection and inventory of burn areas as a result of fire.

In forestry, satellite data are most often used to determine the health of forests for large areas. Satellite data is used due to its availability for large areas and frequent acquisitions but often aerial or UAV imagery is also used for detailed analyses (RGB NIR images and LIDAR data).

Large scale of data, Big Data processing are used more and more often, in particular with machine learning solutions, where the amount of data used to train algorithms is of key importance. Estimating damage to tree stands, inventory of areas affected by fire or flood are relevant examples of how this type of data can be used. However, many of the potential applications require much greater resolution than that offered by the Copernicus Data. In case of hunting damage assessment, detailed forest health assessment or expansion of individual invasive species, a resolution of 15-50 cm is required.

SGIS business products (SprintMAP, ITI portal, GIS-NET.MPZP) are delivered to the client always with intuitive Graphical User Interface, so it would be hard to offer directly CANDELA products. Only very advanced users may be interested in tools that require programming and access from the code level, but CANDELA tools could be used as part of the systems built and delivered to customers by SGIS.

The greatest benefits of CANDELA tools include the possibility of relieving the IT infrastructure (on the client's or SGIS's side). These tools also allow us to compare the results of similar analyses carried out manually (using ArcGIS, QGIS, ENVI) or with the solutions developed by SGIS. The lack of GUI is a limitation as mentioned above.

The greatest potential of CANDELA solutions is related to the possibility of regular, automated monitoring of the dynamics of changes in forest health for large areas. With CANDELA project results SGIS can promote using EO data as input data. However, to properly estimate the impact by CANDELA solutions on forestry business in Poland, more business research needs to be done.

2.3 User capacity impact to exploit the Copernicus Data in general

As EO data are huge, downloading them through the Internet is not an effective way of working. Moreover, it is possible to download only two EO data at once with free account, so the best way to process big amount of EO data is to do it in the same cloud environment where they are hosted.

So, the main advantage of CANDELA platform is that its computing and storage resources are hosted on one of the DIAS cloud. Thus, end users have a direct access to the entire CreoDIAS catalogue though the CreoDIAS library. This library allow end users to make different kind of requests thanks to several parameters, such as the bounding box or the city name for the area of interest, the period of time or the more recent image before or after one event for the date of interest, etc… Once the EO products are selected, a symbolic link to each product is created so data are not duplicated. This process is almost immediate.

For instance in the Natura 2000 use case, whose goal is to monitor the sites of nature preservation in the entire France every month, the first step consists in making a request for the more recent Sentinel-2 products for every tile that represents France at the end of each month (see figure below). This step takes about 10 seconds to process 98 tiles.

Create a folder that contains Sentinel-2 products over the entire France

[2]:
```
import creodias_lib as creodias
from shapely.geometry import Polygon

enddate = '2019-05-01'
result = creodias.search('Sentinel2',
                         productType='L2A',
                         before_date=enddate,
                         time_lapse=28,
                         geometry=Polygon([(2.5, 51.1), (8.3, 49.0), (8.0, 43.7), (3.3, 42.2), (-2.0, 43.3),
                         cloudCover='[0,20]',
                         status='all')

creodias.display_products_identifier(result[0])

folder_link = './links'
#creodias.delete_links_in_folder(folder_link)
creodias.create_links(result[0], destination=folder_link)
```

```
Search for results...
98  products found
```

Product Identifier
/eodata/Sentinel-2/MSI/L2A/2019/04/30/S2B_MSIL2A_20190430T112119_N0211_R037_T30TWT_20190430T130524.SAFE
/eodata/Sentinel-2/MSI/L2A/2019/04/30/S2A_MSIL2A_20190430T103031_N0211_R108_T31TGJ_20190430T140106.SAFE
/eodata/Sentinel-2/MSI/L2A/2019/04/30/S2A_MSIL2A_20190430T103031_N0211_R108_T31UGR_20190430T140106.SAFE

Figure 1: Nature 2000 use case. Sentinel-2 products (first step)

Another advantage of CANDELA platform is that a Kubernetes cluster has been deployed on top of it. Kubernetes acts as an orchestrator and ensures that all the components have access to computation resources and can communicate between each other. This environment is almost mandatory to deal with big data use case.

The second step of the Natura 2000 use case consists in running the optical change detection pipeline in order to compare all the current Sentinel-2 products with the ones of the previous month. This optical change detection tool is a service of the CANDELA platform that is orchestrated by Kubernetes.

This step is divided in three sub-phases:

- ***JP2Tiff:*** convert Sentinel-2 products in geoTiff images with the bands of interest (see figure below).

```
[5]: pipe.JP2Tiff(folder_link, folder_tif, band='02,03,04,08')

[5]: [<owslib.wps.WPSExecution at 0x7f9b9bbe99d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf7710>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf7790>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf79d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf7c90>,
      <owslib.wps.WPSExecution at 0x7f9b9bbe9d90>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf7810>,
      <owslib.wps.WPSExecution at 0x7f9b9bbff750>,
      <owslib.wps.WPSExecution at 0x7f9b9bbfff90>,
      <owslib.wps.WPSExecution at 0x7f9b9bbe97d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bbe9090>,
      <owslib.wps.WPSExecution at 0x7f9b9bbf7250>,
      <owslib.wps.WPSExecution at 0x7f9b9bbfffd0>,
      <owslib.wps.WPSExecution at 0x7f9b9bbff090>,
      <owslib.wps.WPSExecution at 0x7f9b9ab8ab10>,
      <owslib.wps.WPSExecution at 0x7f9ba0bc7d10>
```

Figure 2: Nature 2000 use case. Second step; JP2Tiff (1)

This sub-phase takes a little more than one minute to process 1 tile (see figure below) and about half an hour to process 98 tiles because the environment allows to launch several processes in parallel (here 5 processes at the same time)

```
2020-07-15 21:24:09,077 INFO Kubernetes process id: peg2tiff3processingmichelle20aubrun20200715212408jgnfma4ntcpvuq
started
2020-07-15 21:25:15,684 INFO Processing service launch and execution duration: 66.6061739922s
2020-07-15 21:25:15,728 INFO Pod memory usage (in MB) mean : 1170.6859375 max : 1867.7578125
2020-07-15 21:25:15,728 INFO Pod cpu usage (in core) mean : 2.57580763975 max : 3.257265809
2020-07-15 21:25:15,728 INFO Process success !!
```

Figure 3: Nature 2000 use case. Second step JP2Tiff (2)

- ***ExtractInfo:*** regroup corresponding geoTiff images (see figure below).

```
[6]: pipe.ExtractInfo(folder_tif, folder_tif)

[6]: <owslib.wps.WPSExecution at 0x7f9b9bb1e8d0>
```

This sub-phase is really fast (see figure below).

```
2020-07-15 22:37:08,840 INFO Kubernetes process id: ractinfo3processingmichelle20aubrun20200715223707rwow5sdktz6doq
started
2020-07-15 22:37:15,362 INFO Processing service launch and execution duration: 6.52183103561s
2020-07-15 22:37:15,421 INFO Pod memory usage (in MB) mean : None max : None
2020-07-15 22:37:15,421 INFO Pod cpu usage (in core) mean : None max : None
2020-07-15 22:37:15,421 INFO Process success !!
```

Figure 4: Nature 2000 use case. Second step; ExtractInfo (1)

The result is a json file that contains the different corresponding images (see figure below)

```
▼ root: {} 98 keys
  ▼ 1: {} 1 key
    ▼ Dataset: [] 2 items
        0: "T29UQP_20190420T112119.tif"
        1: "T29UQP_20190505T112121.tif"
  ▼ 2: {} 1 key
    ▼ Dataset: [] 2 items
        0: "T30TUT_20190427T110629.tif"
        1: "T30TUT_20190522T110621.tif"
  ▶ 3: {} 1 key
  ▶ 4: {} 1 key
  ▶ 5: {} 1 key
  ▶ 6: {} 1 key
  ▶ 7: {} 1 key
  ▶ 8: {} 1 key
  ▶ 9: {} 1 key
  ▶ 10: {} 1 key
  ▶ 11: {} 1 key
  ▶ 12: {} 1 key
  ▶ 13: {} 1 key
  ▶ 14: {} 1 key
  ▶ 15: {} 1 key
  ▶ 16: {} 1 key
  ▶ 17: {} 1 key
  ▶ 18: {} 1 key
  ▶ 19: {} 1 key
```

Figure 5: Nature 2000 use case. Second step; ExtractInfo (2)

- ***ChangeDetection:*** compute change detection maps (see figure below).

```
[7]: pipe.ChangeDetection(folder_tif, folder_tif, os.path.join(folder_tif, 'Dataset.json'), './models/dense_4_banc

[7]: [<owslib.wps.WPSExecution at 0x7f9b9bbe99d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bb829d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bb70690>,
      <owslib.wps.WPSExecution at 0x7f9b9bb828d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bb672d0>,
      <owslib.wps.WPSExecution at 0x7f9b9bb82a10>,
      <owslib.wps.WPSExecution at 0x7f9b9bb82a90>,
      <owslib.wps.WPSExecution at 0x7f9b9bb82350>,
      <owslib.wps.WPSExecution at 0x7f9b9bb0b890>,
      <owslib.wps.WPSExecution at 0x7f9b9bb67790>,
      <owslib.wps.WPSExecution at 0x7f9b9bb82790>,
      <owslib.wps.WPSExecution at 0x7f9b9bb0be50>,
      <owslib.wps.WPSExecution at 0x7f9b9bb0b550>,
      <owslib.wps.WPSExecution at 0x7f9b9bb17a90>,
      <owslib.wps.WPSExecution at 0x7f9b9bb82310>,
      <owslib.wps.WPSExecution at 0x7f9b9bb0b110>,
```

Figure 6: Nature 2000 use case. Second step; ChangeDetection (1)

This sub-phase takes a little more than three minutes to process 1 change detection map (see figure below) and a little more than 5 hours to process 98 change detection maps. This service cannot run in parallel because there is only one GPU.

```
2020-07-15 23:14:31,970 INFO Kubernetes process id: etection3processingmichelle20aubrun20200715231429afng5f8lfu0fkq
started
2020-07-15 23:17:44,280 INFO Processing service launch and execution duration: 192.309558153s
2020-07-15 23:17:44,332 INFO Pod memory usage (in MB) mean : 24316.8541667 max : 31626.3945312
2020-07-15 23:17:44,332 INFO Pod cpu usage (in core) mean : 21.5052996744 max : 26.780903543
2020-07-15 23:17:44,333 INFO Process success !!
```

Figure 7: Nature 2000 use case. Second step; ChangeDetection (2)

Thanks to IRIT tool, CANDELA allows users to index EO data with non-EO data and to make requests using multi-criteria based on non-EO data. Moreover, several formats are supported by IRIT tool, such as shapefile and geojson.

The last step of the Natura 2000 use case consists in triplifying the change detection results (based on EO data) according to the Natura 2000 sites (based on non-EO data). Thus, the result provided at pixel level is converted at site of interest level (see figure below).

Triplify change detection result on Natura 2000 geofeatures

```
[1]: import os
import semanticlib

folder_tif = '/home/jovyan/work/tifs'
for filename in os.listdir(folder_tif):
    if filename.startswith('result_'):
        semanticlib.Triplify(feature_file='/home/jovyan/work/Natura2000_fixed.geojson',
                             raster_file=os.path.join(folder_tif, filename),
                             feature_type='Natura_2000',
                             output_folder='/home/jovyan/work/geojson',
                             threshold1=0.2, threshold2=0.4, threshold3=0.6
                            )
```

Figure 8: Nature 2000 use case. Third step (Triplify change detection result on Natura 2000 geofeatures)

This step takes about 45 minutes to run one tile on all the Natura 2000 sites in France (see figure below).

```
2020-10-16 09:09:15,848 INFO Kubernetes process id: ificationprocessingmichelle20aubrun20201016090915zlyeea51vn8fcg
started
2020-10-16 09:50:51,413 INFO Processing service launch and execution duration: 2495.56497979s
2020-10-16 09:50:51,496 INFO Pod memory usage (in MB) mean : 1501.66312123 max : 2100.41796875
2020-10-16 09:50:51,496 INFO Pod cpu usage (in core) mean : 1.00162782847 max : 1.215111136
2020-10-16 09:50:51,497 INFO Process success !!
```

Figure 9: Nature 2000 use case. Third step (Triplify change detection result on Natura 2000 geofeatures)(2)

Once the Natura 2000 use case has finished to run, any operator can monitor the changes that occurred on all the Natura 2000 sites in France during the last month.

This scenario illustrates how big data technologies can be precious for earth observation data processing. This permits to process a large area (entire France here) over a certain period of time. It provides automatic and rapid tools to monitor and analyse large scale of data, which cannot be done manually.

3 Innovative applications creation

3.1 Innovative applications allowed in Agriculture domain

Big data solutions based on Copernicus earth observation product can play an important role by providing innovative large scale applications like health monitoring of crop, crop inventories, monitoring of certain type of crops. It can also allow to develop water related innovative applications such as crop irrigation, drought monitoring or soil moisture and manage its usage. Copernicus imagery combined to Big data on Copernicus earth observation can help to optimize and manage land use.

CANDELA platform will allow developing applications that rely on large-scale processing such as drought monitoring. Indeed, drought monitoring has gained a lot of attention from governments and food agencies due to climate change. It affects the growing of crops, which causes crises in food production. With the capabilities of the CANDELA platform, the development of such an application can be feasible as it is adapted to process large amounts of data.

Indeed, with the continuous increase of the world's population, crop production must be stepped up.

We believe that big data will enable farmers to make smart and informed decisions such as the crop type to grow for better profitability, when to harvest, etc.

This enables them to make smart decisions, such as what crops to plant for better profitability and when to harvest. Additionally, it will allow farmers to follow food agencies and government regulations on the use of fertilizers in crops. Big data will allow farmer to better manage the risks using data-driven farming.

CANDELA platform can permit to monitor a large scale of crops and to record changes over a long period of time. This will help farmers to manage yield mapping and then increase productivity and yield.

Such technologies can also help public authorities for subsidy controls. Automatic and continuous monitoring of the cultivated areas on a large area allows to reduce the need for on-the-spot checks, which are costly and limited in scale.

However, for some specific applications needing precision, such as the vineyard monitoring for example, the resolution offered by the Sentinel earth observation data is not enough. By opening the CANDELA platform to other commercial data provider, more applications could be considered.

3.2 Innovative applications allowed in Forestry domain

Copernicus imagery combined to big data technologies can permit to develop innovative applications to map the forestry resources, such as the inventory of forest resources, monitoring of changes, evaluation of land productivity for forest. It can also allow to monitor the resources and for example manage fires or predict wood provision.

Although the very subject of the analyses carried out by CANDELA tools is not new to forestry in Poland, the solutions provided by CANDELA may help in conducting forest health analyzes on a much larger scale. Most forest analyses in Poland are currently carried out at the level of individual forest districts. Tools that enable data processing for much larger areas can be part of monitoring systems at the voivodeship or country level.

Big data in earth observation is an important step in developing analytical capabilities on the available data. Nowadays, with the free access to data like Copernicus program, the ability and having the appropriate background to use and analyze data is crucial. The analysis of data from many time series, including historical ones, can provide new conclusions about the health of forests, which would not be possible to obtain with analyses conducted in isolated areas and over short periods of time.

Such technologies are a powerful tool for the mapping of the resources. Thanks to earth observation on a large scale, the inventory of forest resources is made easier. Also, by the monitoring of changes in resources over the time, it will help to supervise damages notice and evaluation for example or to detect illegal changes in the land cover and usage in forest areas, such as clear cut or urban area expansion. Nowadays indicators on the forest health are very important for world climate monitoring and these applications are then major to facilitate the task.

Big data in earth observation also helps for resource monitoring. Automatic and continuous monitoring of the forests allows to measure forest health and degradation and then to supervise and predict the fuel wood yield and other resource supplies.

However, the change detection is not so easy with forests and changes do not necessarily reflect variations in the forest. But it may indicate changes in the atmospheric conditions or humidity conditions, it can also be seasonal changes on the vegetation. So, applications developed for forest monitoring should be really robust to ease its use on a large scale and for many users.

3.3 Innovative applications allowed in general

Big data technologies in earth observation business are destined for growth and will be powerful tool for many kinds of applications in the coming years.

As for example with the CANDELA Natura 2000 use case presented in the previous chapter, we can imagine a wide variety of applications, such as the monitoring of the Natura 2000 areas to early detect damages, such as fires, drought or floods, and improve rescue and protection of the critical areas. By monitoring the changes over a long period of time, it can also help to monitor the ecosystem health and evaluate the impact of the climate on the protected ecosystems. Public authorities can also use this kind of tool to control illegal activities disturbing these areas, such as illegal resource exploitation (water or wood for example), illegal constructions, unjustified vehicle traffic, etc.

Copernicus imagery combined to big data technologies has a strong potential of development in all the domains for which the earth observation is suitable: urban monitoring, coastal and marine exploitation and preservation, energy exploitation and implementation, environment monitoring or security operations.

Nowadays, while population is increasing and cities are expanding, urban monitoring is a real concern and expansion needs to be monitored to ensure it proceeds on a sustainable basis. Big data on earth observation represents a key tracking tool to make sure we do not overexploit the environmental resources and we do not worsen the quality of life and safety of urban population.

The technologies developed in the frame of CANDELA allow to monitor the changes to inventory the land cover in cities and peripheral areas in order to monitor urban sprawl and cities progress in making conservation efforts. By making easier the monitoring on a large period of time and a large scale areas, we can imagine to map and monitor large areas in order to assess the urban adaptation to climate change and to analyse the environmental and socio-economical evolutions.

In this context, the technologies developed in the frame of CANDELA are a good start to imagine new applications for urban planning activities. We can actually imagine that thanks to monitoring and change detection on earth observation data, we can map green areas in and around cities, and analyse different determining factors such as density of buildings, green areas location, repartition of air temperature, etc., to better understand how to design smarter cities.

Earth observation big data can support lots of marine and maritime applications. These technologies, such as CANDELA platform, allow to preserve the littoral by monitoring and prevention of coastal erosion or by forecasting sea level rising and storm surges. It is also a mean to monitor marine and coastal protected areas in order to improve their protection and to anticipate potential deterioration.

Combined with field data, it can help fishing industry to map fishing zones or to forecast alga blooms, in order to ensure a sustainable fishing, to provide fishing trends and statistics or to limit the waste of contaminated shellfish.

The technologies developed in the frame of CANDELA permits to monitor the oceans and to forecast new areas accessible due to ice melting in order to optimize ship navigation routes.

It can also be a support to development of marine renewable energies thanks to a mapping of the foreseen areas and a better knowledge of any event activities happening on these areas.

Big data technologies in earth observation can be a decisive support for the growing market of renewable energies. For solar energy for example it can helps to monitor many foreseen sites and to determine the irradiance conditions, the presence of dust and the sky conditions (if it is clear the majority of the time) on each site. This will help to select the more appropriate site and to support efficiency and production yield analysis. For the hydropower business, the earth observation data can be used to analyse land cover and land changes or to estimate the volume and distribution of snow in order to provide information for plants management.

The technologies developed in the frame of CANDELA can support oil and gas activities by detecting oil seeps and creating topographical maps in order to support exploration and drilling activities. It can also help oil and gas exploitation activities by detecting oil spills and leaks, by monitoring extreme weather conditions or by detecting hazards to workers and the platforms (for example ice floes).

CANDELA platform or such big data platform for earth observation data can be a precious help for environment monitoring and in particular natural disaster. There are more and more extreme events (fires, floods, drought, storms, etc.), the trend is increasing, so this kind of tool will contribute to emergency responses but also to risks forecasts and early warning operations.

Risks modelling can be improved thanks to the analysis of previous natural disaster observations. In this context, CANDELA platform, which goal is to ease the analysis and the change detection over a long period of time, can help to understand the phenomenon and calibrate the forecast models.

With high-performance models, big data technologies on earth observation can permit to continuously monitor the indicators on a large scale to perform a short term and long term forecast. This will allow to early warn and alert authorities.

At the end, if the natural disaster happens, these technologies will support the crisis management by providing rapid mapping of disasters and a comprehensive vision of the affected area. This will help to optimize and organize the workforce.

Big data technologies in earth observation can be a strategic stand for security and safety operations taking place both on land and on the seas.

Earth observation data provides access to an exhaustive understanding of the situation. And combined with big data technologies, it permits to monitor a large coverage (borders or specific maritime zones) and continuously analyse and track the detection of abnormal behaviours and illicit activities (by non-reporting vessels, illegal fishing activities or illegal immigration)

Thanks to global monitoring over a long period of time, we can also imagine applications providing vulnerability assessment and risk analysis on land borders, which is useful for intelligence needs.

As seen before, these technologies are also a crucial support for oil spill detection, monitoring and mitigation.

Maritime Search and Rescue operations is a strategic activity in Europe and these activities can be hold up by the monitoring of the traffic and, in polar regions, of icebergs.

4 Economic value generation

4.1 Economic value generation in Agriculture domain

By using these big earth observation data applications, we can imagine that farmers will be able to increase productivity by optimizing the use of water, fertilizers, seeds and pesticides, leading to improved profitability and cost efficiency. By reducing the amount of inputs used, the farmer can also help protect the environment and reduce land, water and air pollution.

The monitoring of crops can have also a variety of impacts. Land use classification can help to assess crop location changes and see if there are any trouble changes or trends.

Monitoring can also be used to support subsidy control, thereby reducing the need for on-the-spot checks, which are both costly and limited in scale. Crop monitoring can help identify environmental factors that need to be taken into consideration.

Data such as wetness indicators can help identify the level of irrigation needed in regions. Monitoring the crops and surrounding land will also indicate if there is unhealthy vegetation and a lack of water, or if there is an increase in algae in the water, which could indicate fertilizer pollution.

Integrating the CANDELA platform in TerraNIS' offers would allow to have economic benefits:

- Automate resource consuming data manipulation processes such as image retrieval and image pre-processing. This would allow to reduce production costs as it saves both the required time and costs.
- Propose new services based on large scale processing such as drought monitoring.
- Provide these services to new customers such as food agencies and governmental bodies.
- Scale existing services to adapt for the increased number of clients due to the capabilities of the platform to scale when needed.
- The acquired experience during the CANDELA project would allow TerraNIS to participate in other research and development projects, which achieves an added-value for TerraNIS.

The services that would be provided based on the CANDELA platform and analytics tools could achieve the following social impacts on the agriculture domain:

- Helping to tackle the challenge of food supplies for the growing population. This can help policymakers to take informed decisions based on patterns found on the large amount of earth observation data.
- Allowing the farmers to achieve better profitability by helping them to select the crop type to grow. Moreover, the tools that allow monitoring the agricultural fields which allow farmers to fight against insects and anomalies by taking early decisions and hence avoiding economical loss caused by the damage of their crops.
- Providing tools for farmers that allow them to manage the fertilizers and respect the regulations and avoid sanctions in case they do not follow the regulations.

The platform and analytics tools can achieve many environmental impacts in the field of agriculture. For example:

- It allows farmers to better manage the fertilization process as it has a direct impact on the soil pollution.
- Additionally, the excess use of fertilization on crops used for food production has a negative impact on human health. Controlling their doses is of priority in this case.

- With some adaptation, the proposed tools can help farmers in irrigation management. Water is a very important natural source and it has become one of the main interests of governments and agricultural associations. This will help in economizing water consumption.
- Monitoring tools are adapted to ensure green societies, which is directly correlated with the population well-being.

4.2 Economic value generation in Forestry domain

Thanks to big data solutions based on Copernicus earth observation, we can imagine that forest authorities will be own accurate forest mapping, which is extremely useful for storm damages assessment for example, inventory, or validation of forest strand for wood purchasers.

Bid data on earth observation is a fundamental tool for deriving statistics on deforestation and land use change and is critical for sustainable development at local, national, regional and global scales. Forest monitoring applications will help forest management stakeholders to create custom maps, observe forest trends or receive alerts on their area or the entire world.

Integrating the CANDELA platform in SmallGIS offers would allow to have economic benefits:

- Expand the company offer and providing new services for our current customers.
- Propose solutions leading to automation of large scale processes. Not requiring the customer to have high hardware and technical background.
- Opportunity to participate in new projects, based on the experience gained in the CANDELA Project.

The development of tools and applications in the field of forestry including CANDELA Platform support the following societal impacts:

- Forest ecosystem preservation and monitoring in urban areas leading to
  - the normalization of temperature and humidity distribution in the city,
  - enhancing biodiversity,
  - pollution reduction,
  - real estate values increase.
- Preservation of natural forests for leisure and recreation.

Forests are one of the most crucial components of the environment. The values in the CANDELA project allow for the following environmental impacts:

- The ability to monitor large areas of forests on an on-going basis, leading to a continuous improvement of knowledge on the impact of various factors on their well-being and the assessment of protection measures applied.
- Support forest maintenance and preservation for climate mitigation.

4.3 Economic value generation in general

All these new technologies and the range of applications that are made possible largely contribute to numerous economic, societal and environmental impacts. From the new applications presented in the previous chapter, here are estimated the potential impacts they can provide.

The figures present the domain, applications made possible and their benefits.

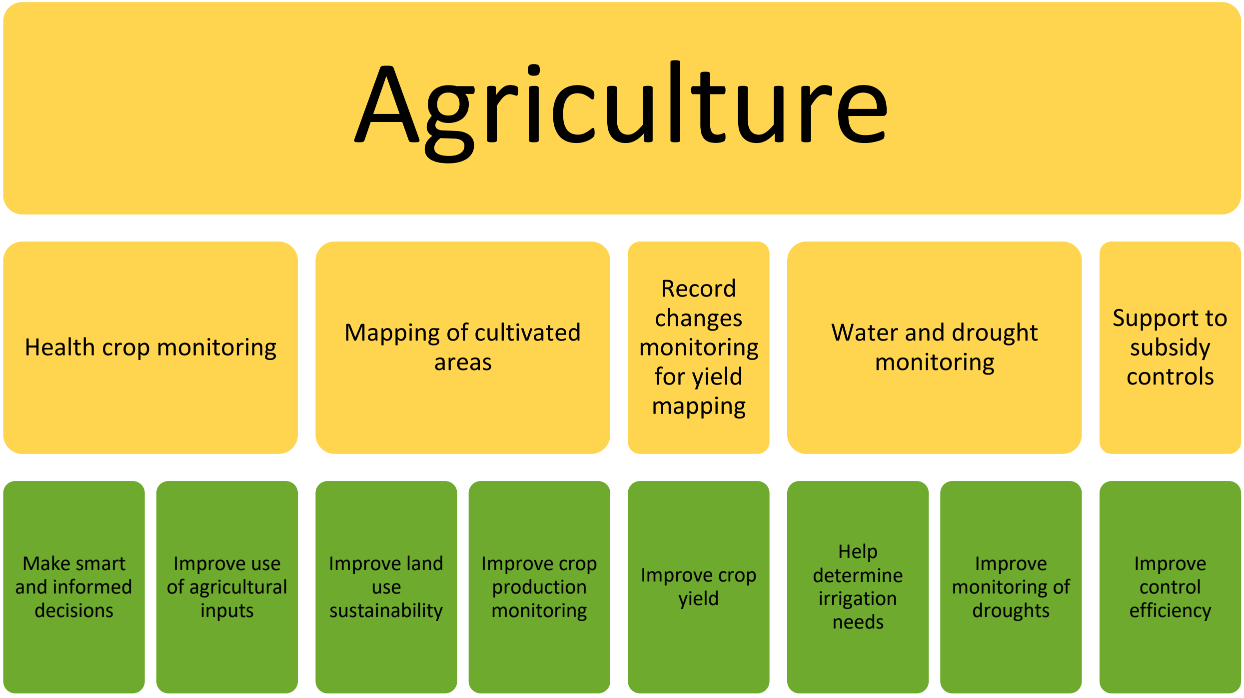

Figure 10: Economic value generation (Agriculture)

Big data solutions based on Copernicus earth observation product in the field of agriculture allow the following economic benefits:

- Cost savings and higher profits thanks to more efficient production and limited inputs
- Market stability
- Increase crop production by the monitoring of yield and constant progress
- Cost savings for authorities from improved efficiency on the controls (limited needs to on-the-spot checks)

The development of tools and applications in the field of agriculture supports the following societal impacts:

- Improved safety to human health by limiting the land and water pollution
- Reduce risk to food insecurity for human by monitoring on a large scale the world production efficiency
- Ensures the implementation of EU policies

It permits also the following environmental impacts:

- Decreased pollution to air water and soil
- Protection of biodiversity
- Reduce water consumption which is a limited resource on certain areas

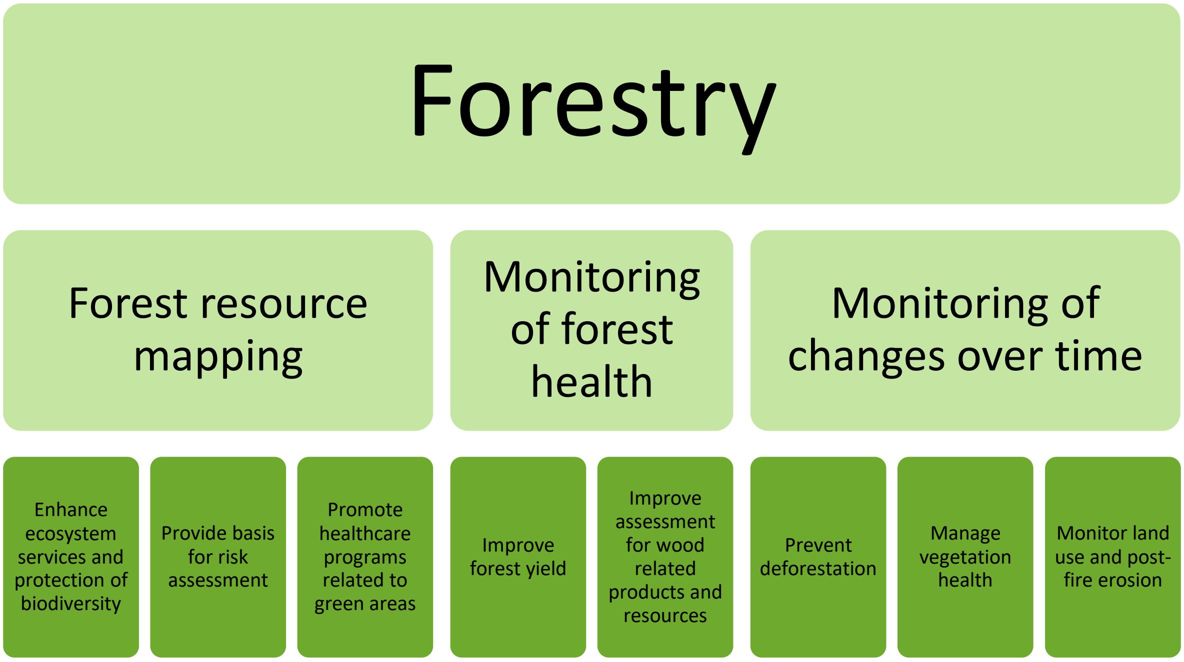

Figure 11: Economic value generation (Forestry)

Big data solutions based on Copernicus earth observation product in the field of forestry allow the following economic benefits:

- Cost savings from risk reduction on forest resource
- Cost savings and higher profits thanks to improved yield
- Market stability thanks to a better evaluation of wood related resources
- Reduce economic damage linked to deforestation

The development of tools and applications in the field of forestry supports the following societal impacts:

- Enhancement of quality of life for people by promoting healthcare programs on green areas
- Sustainability of resources exploitation

It permits also the following environmental impacts:

- Enhancement of ecosystems protection
- Environmental forest protection
- Enhancement of resource restoration by a comprehensive knowledge of land use and fire impacts
- Enhancement of resource recovery by improving the planning and intervention on forest disasters
- Climate mitigation on the vegetation health

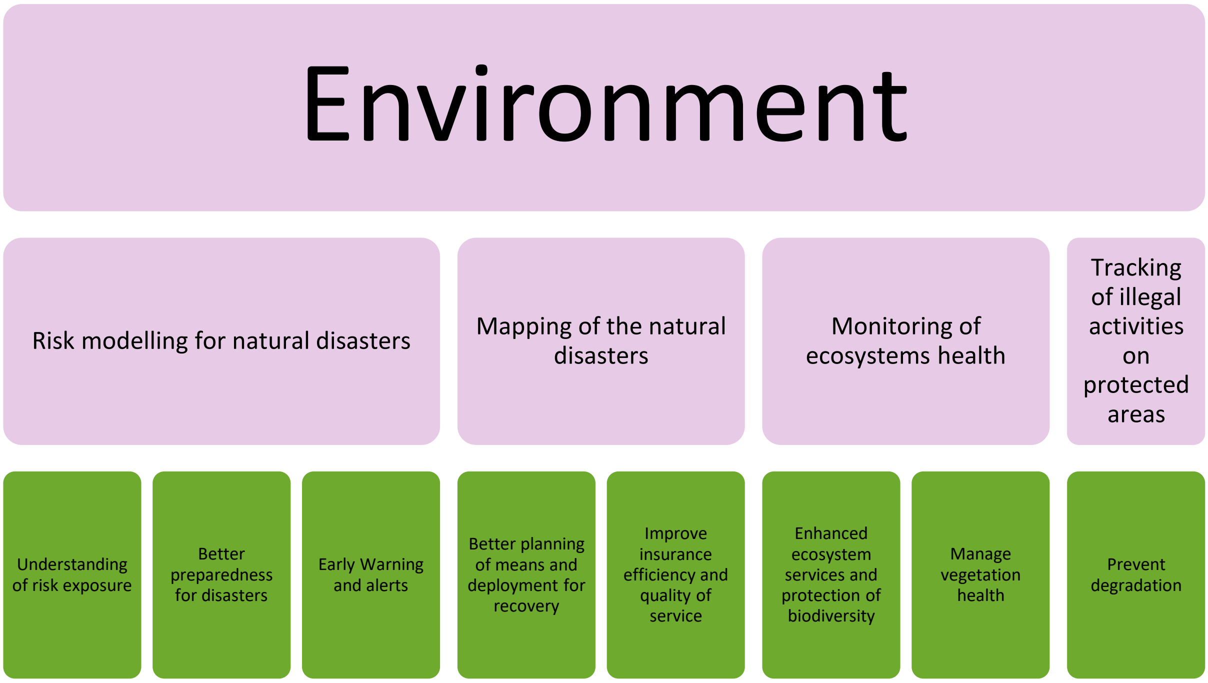

Figure 12: Economic value generation (Environment)

Big data solutions based on Copernicus earth observation product in the field of environment allow the following economic benefits:

- Lower financial risk due to natural disasters
- Reduction of economic damages thanks to risk prediction and early warning
- Reduction of material losses by improving the reactivity and quality of recovery
- Cost savings by providing appropriate financial protection scheme
- Cost reduction from risk reduction and response improvement to natural disasters
- Cost savings from improved efficiency on vegetation health protection
- Better assessment of damaged areas
- Reduction of economic damage caused by illegal exploitation of the protected resources

The development of tools and applications in the field of environment supports the following societal impacts:

- Better planning of field interventions
- Contribution of Europe on the global initiatives for civil protection
- Reduction of precariousness by improving timeliness for insurance handling process

It permits also the following environmental impacts:

- Reduction of environmental damages thanks to risk prediction and early warning
- Customer satisfaction improvement thanks to more targeted actions of insurance companies
- Enhancement of ecosystems protection
- Climate mitigation on the vegetation health
- Enhancement of the resource protection of natural areas

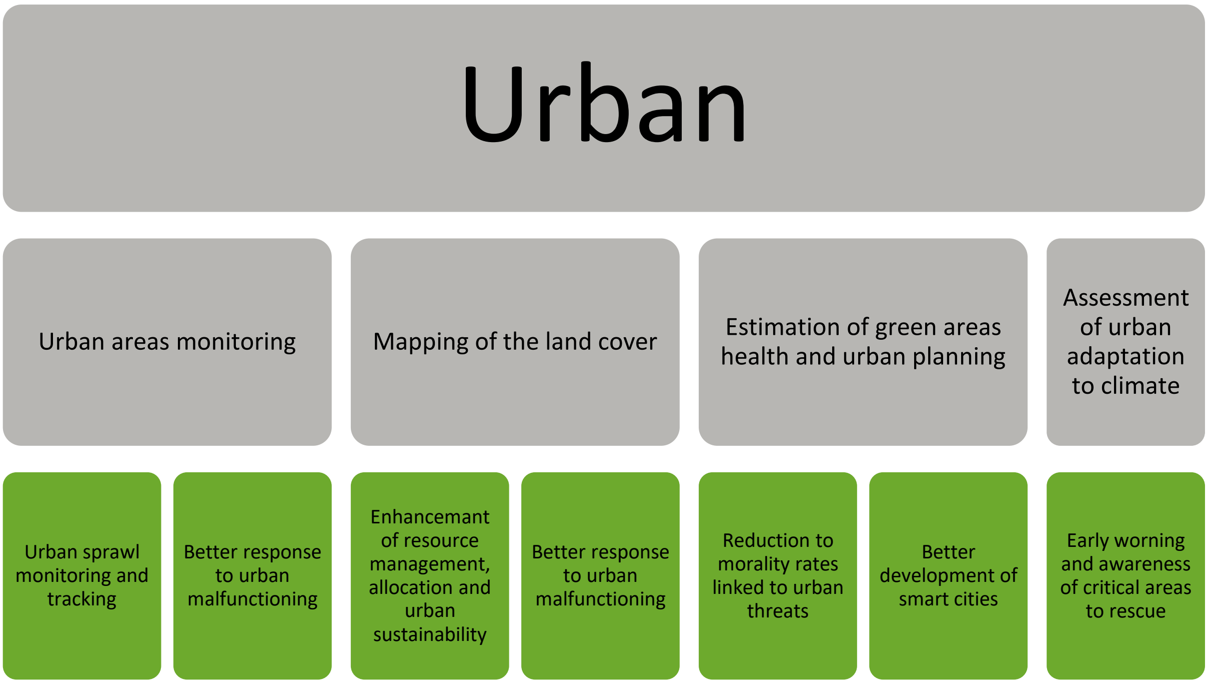

Figure 13: Economic value generation (Urban)

Big data solutions based on Copernicus earth observation product in the field of urban allow the following economic benefits:

- Enhancement of urban and green areas resource allocation
- Reduce risk and damages on critical areas to rescue due to climate change impact
- Reduce damage on critical areas to rescue due to climate change impact

The development of tools and applications in the field of urban supports the following societal impacts:

- Enhancement of urban sustainability
- Damage and mortality risk limitation

It permits also the following environmental impacts:

- Reduction of environmental pressure
- Enhancement of urban ecosystem recovery
- Enhancement of human environmental impact

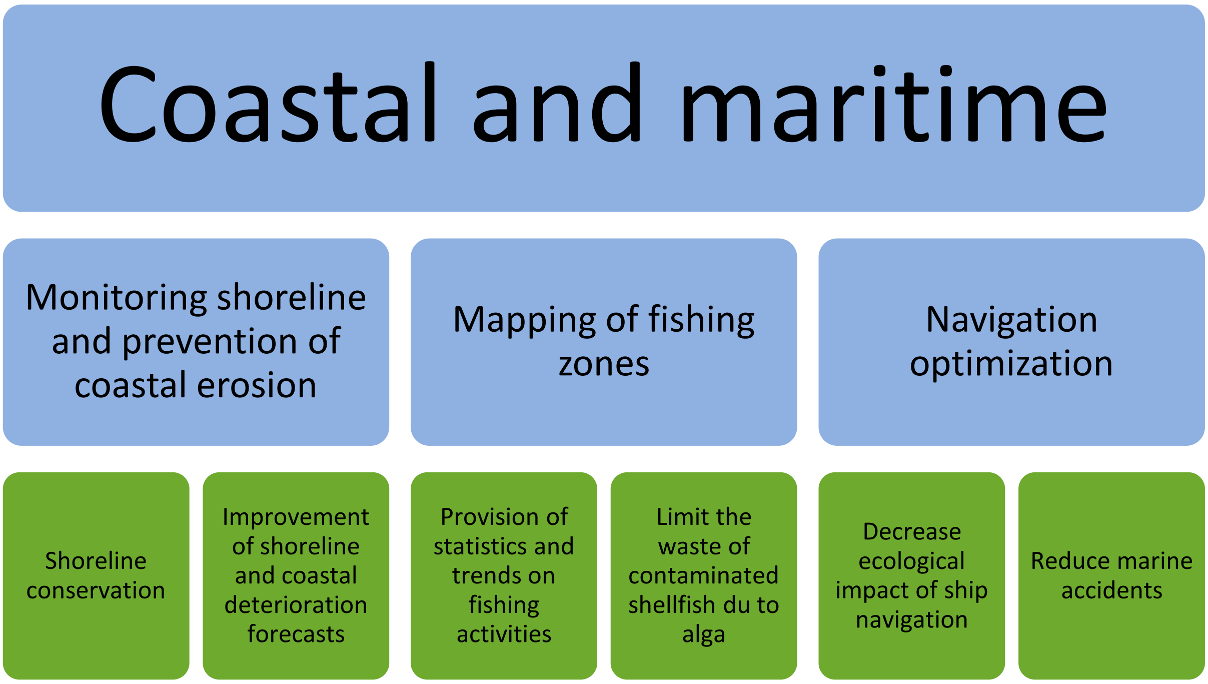

Figure 14: Economic value generation (Coastal and maritime)

Big data solutions based on Copernicus earth observation product in the field of coastal and maritime domain allow the following economic benefits:

- Better productivity thanks to limited contamination by alga
- Less spending on fuel for ship navigation
- Reduction of risk and damages due to marine accidents
- Improvement of the attractiveness of coastal areas thanks to shoreline conservation
- Reduction of risk and damages on coastal areas

The development of coastal and maritime tools and applications supports the following societal impacts:

- Design fishery policies and track if regulations are respected
- Mortality risk limitation due to marine accidents
- Improvement of the attractiveness of coastal areas thanks to shoreline conservation

It permits also the following environmental impacts:

- Support sustainable fishing limited to fishing areas
- Reduction of greenhouse gas emissions due to ship navigation
- Shoreline environmental protection

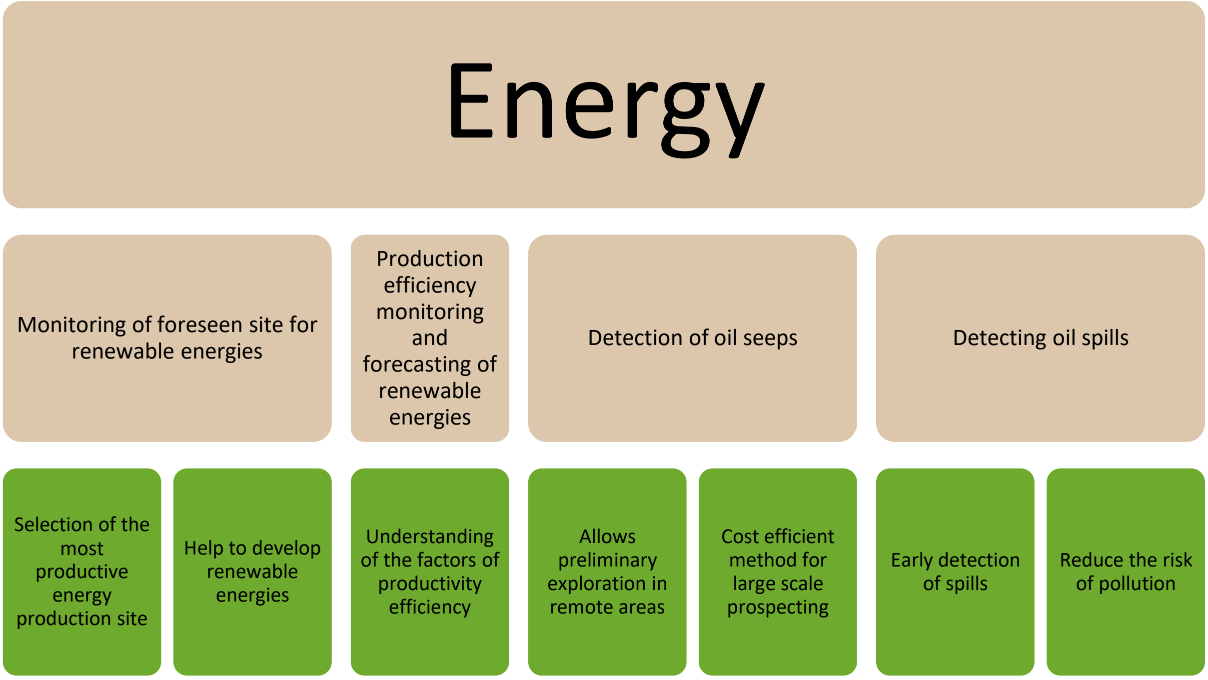

Figure 15: Economic value generation (Energy)

Big data solutions based on Copernicus earth observation product in the field of energy domain allow the following economic benefits:

- Better return on investments for renewable energy facilities development
- Better quality of renewable power production forecasts
- Gain in competitiveness
- Cost savings thanks to early spills detection
- Reduce risk of damage on oil and gas infrastructure

The development of energy tools and applications supports the following societal impacts:

- Lower the risks of nuclear accidents
- Improvement of security of energy supply
- Protection of workers' safety on remote platform
- Reduction of the risk to human health

It permits also the following environmental impacts:

- Improvement of environmental objectives
- Reduction of pollution due to greenhouse gases or nuclear waste
- Protection of the biodiversity and ecosystems

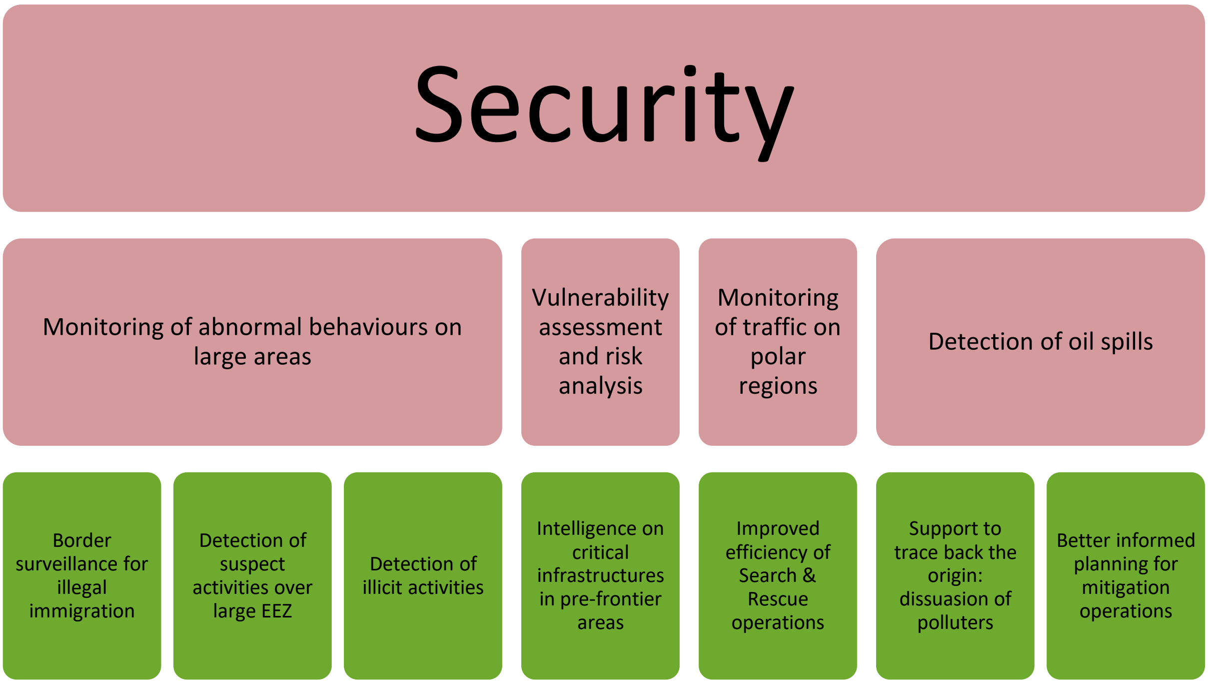

Figure 16: Economic value generation (Security)

Big data solutions based on Copernicus earth observation product in the field of security domain allow the following economic benefits:

- Improvement on fish industry revenues
- Value of illegal goods seizures (drugs or counterfeit goods)
- Reduction of drugs treatment costs
- Reduction of material losses thanks to a better search and rescue operations
- Reduction of economic damages due to oil spills
- Decreased tourism losses due to oil spills
- Reduction of fishing industry losses due to oil spills
- Reduction of cost of cleaning due to oil spills

The development of security tools and applications supports the following societal impacts:

- Better preparedness to conflicts thanks to an improved vulnerability and risk assessment
- Reduction of illegal immigration
- Support to evacuation plans
- Support to humanitarian response
- Reduction of casualties among refugees
- Reduction of casualties in regular traffic
- Mortality risk limitation due to drugs consumption

It permits also the following environmental impacts:

- Enhancement of fish resource sustainability
- Preservation of marine ecosystems
- Reduction of environmental damages due to oil spills

5 Conclusions

Big data technologies with Copernicus earth observation are able to offer powerful applications ensuring good quality, reliability, resolution and revisit.

The automatic data collection and changes detection algorithms combined to additional data developed in the frame of CANDELA provides important positive impacts on the operational use of earth observation data. It permits to save time compared to manual operations, and to provide exhaustive results on the comparison. This solution also permits to filter data to intelligently focus on customer interest. The automation and big data technologies treatments allows to concentrate the operations on specific alarms and to focus on the useful information.

Using big data technologies facilitate to improve the processing time and to process huge amount of data for and extended monitoring, on a large scale. In addition to that, it permits to perform a continuous background monitoring and to alert the operators only when it is needed.

Using Copernicus data is an advantage because the data is free so it's a non-negligible economic impact for new businesses. It also allows to provide 100% European solutions. The acquisition of data is insured by the Copernicus program and then data acquisition and availability are not under customer or user responsibility. Moreover, the data are always updated with a revisit of about 5 days, so the data are automatically available and fresh, allowing lot of applications.

The implementation on DIAS makes these solutions much easier to access and process.

The combination of optical and radar data is very useful for many applications and domains. The complementarity between both optical and radar data offer lot of opportunities and a large panel of potential applications. This potential is also greatly improved by the possibility of combining with additional data.

However, for some specific applications, the resolution needed is not sufficient and may be combined with others commercial data, for complex urban land mapping for example. The developed solution could then be improved by the possibility to include commercial observation data with better resolution.

For some applications, earth observation data only is not enough and may need to be combined with existing data (from on-the-spot visit or environment data for example).

The access to these kinds of platform and its use in not intuitive and is a brake for some users not used with computer technologies. Adding user-friendly GUI and facilitate the access to the platform can also improve the tools more attractive.

CANDELA is focused on optical and radar earth observation data. The integration of Sentinel-3 data by providing altimetry data, sea surface temperature, etc. could improve the possibility of applications (more application in the marine domain for example).

For natural disaster, the delay between the catastrophe and the availability of the images is crucial. Therefore, for quick events such as heavy rains, the uncertainty on the response timeliness limit the interest in Copernicus.

To improve the CANDELA solution, more artificial intelligence could also be integrated to improve the characterization of the changes and the areas mapping.

References

[1] PWC Copernicus Market report, Issue 2, February 2019. Prepared by PwC

[2] Copernicus Open Access Hub (previously known as Sentinels Scientific Data Hub). https://scihub.copernicus.eu/

[3] PEPS. Operating platform Sentinel products (CNES). https://peps.cnes.fr/