Climate Change Initiative Extension (CCI+) Phase 1

New Essential Climate Variables (NEW ECVS)

High Resolution Land Cover ECV (HR_LandCover_cci)

System Requirement Document

(SRD)

Prepared by:

Università degli Studi di Trento
Fondazione Bruno Kessler
Università degli Studi di Pavia
Università degli Studi di Genova
Université Catholique de Louvain
Politecnico di Milano
Université de Versailles Saint Quentin
CREAF
e-GEOS s.p.a.
Planetek Italia
GeoVille

Changelog

Issue	Changes	Date
0.1	First version.	14/05/2019
0.2	Second version including contribution on Cloud Requirements and Data requirements	26/06/2019
1.0	Final version including the overall requirements analysis and introductory chapters	28/06/2019
1.1	Removed references to EEA layers	25/09/2019
2.0	• Modified input Sentinel 1 from SLC to GRD • Modified the year for the static map from 2018 to 2019 • Removed unnecessary input/ancillary data: historical SAR data, Sentinel 1 SLC • Modified the pre-processing chains to adhere to final input data	16/04/2020
2.1	Final version after modification required by the RID	04/05/2020

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change
1.1	FR-01	EEA layers should not be considered as they refer to European area which is not overlapping with CCI+ HRLC identified area of interest	4.3.1	Removed references to EEA layers
2.1	FR-01	Issue and date are wrong (in the header)	N/A	Header and issue have been updated to v2.1 (after applying the corrections required by the RIDs) and delivery date to 04/05/2020
2.1	FR-02	"The CCI HRLC PLATFORM will be deployed on a DIAS environment facility as preference expressed by ESA. In the case of very strict constraints, other options will be evaluated." This sentence disagrees with what was said during the PM #4 at end of January 2020 "It is agreed that the processing will be performed on commercial cloud platforms given that the cost of DIAS is not affordable with the available budget."	2.3 Deployment overview	The paragraph has been updated accordingly. The decision about using a Commercial Cloud (AWS) has been reported and the description of the analysis activity has been updated.
2.1	FR-03	The year for the static map is 2019 and not as reported in the sentence "The Sentinel 2 data for the AOIs in 2018 shall amount to approximately 550k…"	4.5.3.1 Sen2Cor	The paragraph has been updated to describe estimation of resources needed to process with sen2cor a generic year.

Contents

1 Introduction

1.1 Executive summary

Following the activities of user requirements updating according to Climate User Community and other users' consultations, the Consortium has defined the related HRLC products requirements accounting for technical constraints such as main data sources available, spatial and temporal coverage, software and tools for quality control.

This High Resolution Land Cover (HRLC) System Requirements Document defines a set of requirements for an operational system for the European Space Agency (ESA) Climate Change Initiative (CCI+). The system will be used to generate and update yearly the HRLC product in selected areas for the purpose of global climate change and in addition for the analysis at regional level.

The system requirements are derived from input requirements - mainly the Statement of Work and the User Requirements document. As a result, the document defines a set of system requirements for production and reprocessing of large volumes of data as dealing with HR data up to 10 meter resolution. Consequently, requirements cover also performance issues and a preliminary set of requirements for the selection of a Cloud environment dedicated to EO data processing.

The first part of the document presents a synthesis of the concepts and assumptions related to the PLATFORM development for the generation of the High Resolution Land Cover layer.

The first part ends with the analysis of use cases derived by the analysis of Science Team work on Product Specification and on pipeline definition and by the definition of example scenarios on the basis of previous experience by the Engineering Team.

Finally, the requirements are drafted grouped by functional domain to drive following phases of platform development starting with the specification of the platform.

For what concern the environment in which the platform will be deployed, DIAS have been considered as requested by ESA. At the time of writing it is impossible to make the final decision as the DIAS have not responded to the specific questions and also because the correct time to make the decision is actually the design time. Nevertheless, a set of minimum requirements have been identified and a set of preferred requirements have been identified. This analysis will be useful during the design phase to select the best candidate among the DIAS environment.

1.2 Purpose and scope

The HRLC System Requirements Document defines the requirements for an operational system for production of HRLC maps over the three areas selected in this phase.

Input to this document are the Tender Specification and the other Applicable Documents.

- To generate the products specified in the HRLC Product Specification Document (PSD) using data specified in the Data Access Requirement Document (DARD);
- To apply the algorithms specified in the HRLC Algorithm Theoretical Basis Documents (ATBD) for this purpose (not yet finalized and with pending decision at the time of writing);
- To make the product accessible to users as specified in the HRLC User Requirements Document (URD).

Despite, these inputs provide little information on how the HRLC platform should be used, the approach used for this document is to develop various example and representative use case and scenarios for the processing workflow in order to have a starting point for the derivation of requirements. As far as possible, the requirements are derived from inputs, scenarios and the processing workflow.

1.3 Applicable documents

Ref.	Title, Issue/Rev, Date, ID
[AD1]	CCI HR Technical Proposal, v1.1, 16/03/2018
[AD2]	CCI Extension (CCI+) Phase 1 – New ECVs – Statement of Work, v1.3, 22/08/2017, ESA-CCI-PRGM-EOPS-SW-17-0032
[AD3]	Data Standards Requirements for CCI Data Producers, v2.0, 17/09/2018, CCI-PRGM-EOPS-TN-13-0009
[AD4]	User Requirements Document, v1.1, 12/04/2019, CCI_HRLC_Ph1-D1.1_URD
[AD5]	Product Specification Document, v1.0, CCI_HRLC_Ph1-PSD
[AD6]	Data Access Requirement Document v1.0, CCI_HRLC_Ph1-DARD

1.4 Reference documents

Ref.	Title, Issue/Rev, Date, ID
[RD1]	The Global Climate Observing System: Implementation Needs, 01/10/2016, GCOS-200

1.5 Acronyms and abbreviations

API	Application Programming Interface
AOI	Area Of Interest
ARD	Analysis Ready Data
BOA	Bottom Of Atmosphere
CCI	Climate Change Initiative
CRC	Climate Research Community
CMUG	Climate Modelling User Group
CREAF	Centre de Recerca Ecològica i Aplicacions Forestals
ECV	Essential Climate Variables
ESM	Earth System Models
EVI	Enhanced Vegetation Index
FTP	File Transfer Protocol
GCOS	Global Climate Observing System
GDPR	General Data Protection Regulation
GIS	Geographical Information System
HR	High Resolution
IaaS	Infrastructure as a Service
L1C	Level-1C
L2A	Level-2A
LAI	Leaf Area Index
LaSRC	Landsat Surface Reflectance Code
LC	Land Cover
LCC	Land Cover Change
LCCS	Land Cover Coverage Classification System
LCML	Land Cover Meta Language

LCZ	Local Climate Zone
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
MR	Medium Resolution
NDVI	Normalized Difference Vegetation Index
OGC	Open Geospatial Consortium
OWS	OGC Web Services
PFT	Plant Functional Type
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SFT	Surface Functional Type
TOA	Top Of Atmosphere
URD	User Requirements Document
VM	Virtual meeting
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package

2 High resolution land cover platform concepts and assumptions

2.1 HRLC platform overview

The HRLC platform prototype (named PLATFORM in the requirement chapters) scope is to bring the results of the research activities to a pre-operational level by scaling up the processing capacity in order to allow the production of massive land cover mosaics following GCOS requirements.

The drivers for the Phase 1 Production of the project will be:

- To be ready for massive production using scalable processing facilities (e.g. DIAS, Cloud) while maintaining flexibility of the process (e.g. to ease the future enhancement of workflows/algorithms)
- Optimize resource consumption by considering on-demand processing resources (e.g. IaaS, PaaS)
- To be as simple as possible considering the massive production required and the complexity of the overall processing/analysis task
- To be able to manage data from different sources, also dealing with large data volume [TR-12]

The following scheme shows the main process which will be supported by the overall HRLC platform prototype development and the organization within the consortium. In particular, the CCI prototype platform will put the basis for future enhancement by allowing easy link between the research activity and the production activity.

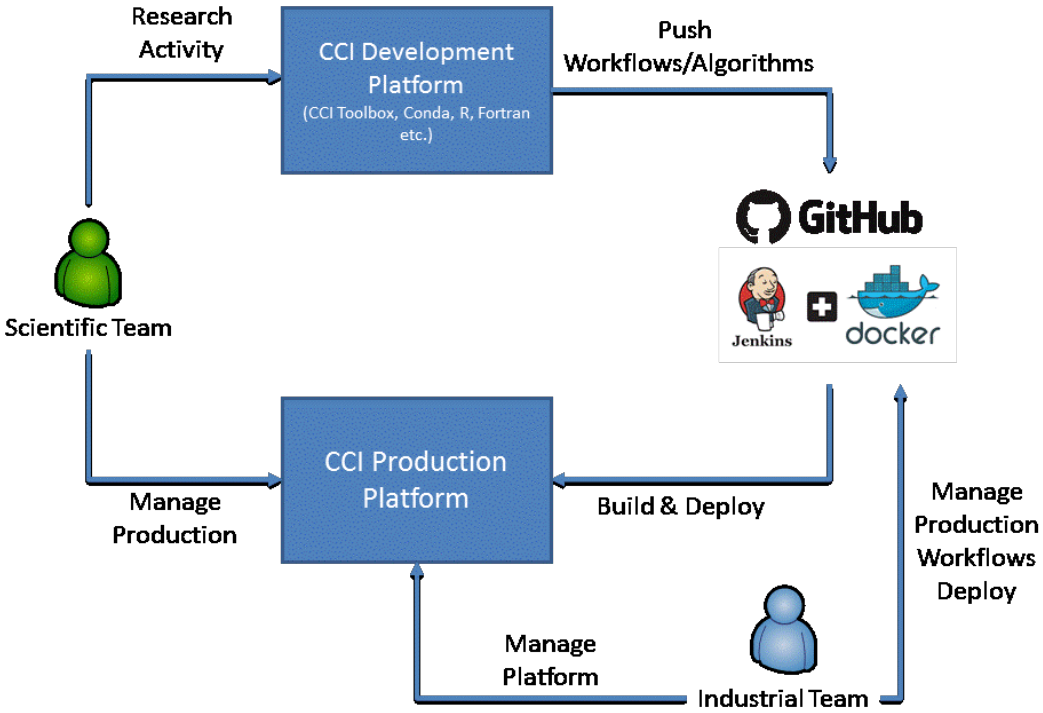

Figure 1. CCI Development/Production Platform Concept

The figure highlights the overall process starting from the research activity using for example the Conda environment (based on the Conda Python-based package manager, https://conda.io/ largely used in the scientific community and used also for the development of the CCI Toolbox). The following step is the push of workflows and algorithms (and test data) into the Github project repository that automatically trigger the generation/test of packages and the final creation of Docker images.

The platform will guarantee the processing over the selected test area starting from the data retrieval/ingestion until the final product generation/storage/conversion/distribution in a systematic way by also optimizing new data ingestion from multiple sources. [TR-15]

2.2 HRLC platform **architecture overview**

The following figure shows a detail of the architecture of the HRLC platform:

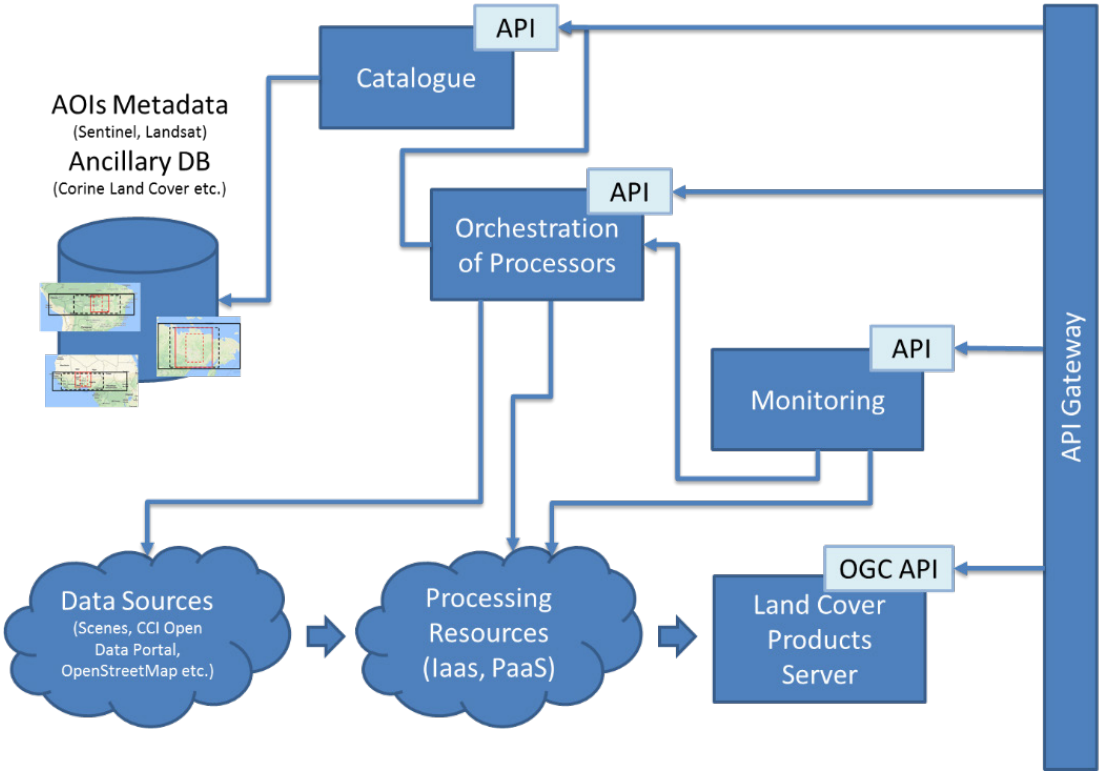

Figure 2. CCI Production Platform: detail on the architecture for the orchestration of massive processing

The internal elements of the architecture are:

- **Metadata and Ancillary DB:** which is the store of metadata information (scene boundaries, cloud cover etc.) and of the ancillary data used by the algorithms. The **Catalogue** API allows to search and download the data using standards like OpenSearch
- **Catalogue**: which is the central point for the discovery of all data hosted and referenced in the platform. For example, in the case of imagery (Landsat, Sentinel), it will be possible to do various search and optimization before accessing the **Data Sources** remotely hosted (e.g. in the original archive such as ESA Scihub, DIAS archives or public cloud resources) [TR-17]
- **Orchestration**: is the central element of the architecture. It handles the requests for processing coming from the API and dispatch the processing jobs by retrieving necessary **Data Sources** and allocating the **Processing Resources** to manage (e.g. execute/modify) the workflow used in the generation the final Land Cover Products. The **Orchestration** has also the objective of optimizing the **Processing Resources** in terms of utilization and of maintaining the necessary information (like steps, version control of algorithms/workflows, input/output data) for rerun of the processing.
- **Monitoring**: is the element devoted to check the status of the processing through the access to the **Orchestration** API in order to provide notification in case of anomalies, errors or even in case of job completion. The Monitoring element also interfaces the Processing Resources to check that the used IaaS/PaaS resources match with the expected
- **HR Land Cover Products Server**: is the element that hosts the final products in an optimized format published as OGC service like WMS and WCS. This server will also provide the necessary functions for the analytical Web Interface
- **Web Interface**: will provide a mapping interface capable of interacting with the WMS server and configuring maps and mashups
- **API Gateway**: is the element used to collect the API in a single reference point to gather metrics of the usage of the platform and to manage the access, authentication and authorization functionalities

In addition to the internal elements of the architecture, the external resources are:

- **Data Sources:** all the external available data sources like:
  - **Imagery Data Sources** like Sentinel 1 and 2 archives hosted by ESA, by the DIAS and eventually cloud providers (e.g. Amazon, Google), like Landsat 4-5-7-8 hosted by USGS and eventually cloud providers (e.g. Amazon, Google), like Service providers and ESA DataWareHouse for the access to VHR imagery [TR-15]
  - **Reference Data Sources** like the CCI Open Data Portal and other external data sources (not stored on the platform). An example of this is OpenStreetMap database.
- **Processing Resources:** all the IaaS/PaaS resources available on platforms like the DIAS and cloud providers

In addition, the **orchestrator** will take care of the **dynamic allocation of resources** according to the characteristics and the needs of each task in the pipelines as follow:

- Virtual Machine instance type: number of vCores, GB of vRAM
- Local storage: GB for performing the processing
- Bucket storage: Access to the long term Object Storage such as S3 (e.g. both Open Telekom Cloud and Amazon support the S3 protocol)

When the pipeline tasks are completed, final results are stored in a S3 bucket using the COG format (Cloud Optimized Geotiff) and can be accessed using S3 protocol implemented in GDAL for inspection, validation and further processing.

Docker will be used mainly to guarantee that pipelines/tasks creation and execution can be managed without loss of time in the installation and compatibility checks.

2.3 Deployment overview

The CCI HRLC PLATFORM will be deployed on a commercial cloud and in particular on AWS (Amazon Web Services). After an extensive analysis of the 5 DIAS it has been concluded that prices are not in the range required for the production as all DIAS are missing an offer based on spare resources like AWS Spot or Google Pre-emptible resources.

During the drafting of the document and the analysis phase the following activities have been carried out: all of them have been contacted for obtaining a feedback specific to the CCI HRLC case on a tailored questionnaire. In this SRD, the requirements have been defined as minimum set of requirements for the deployment of the PLATFORM and execution of pipelines with respect to:

- Availability of IaaS services, the DIAS shall make available a minimum set of IaaS services through API
- Availability of DIAS services:
  - The DIAS shall make available a minimum set of Web Services such as Catalogue, Data Repository
  - The DIAS optionally can provide value added services and data such as Cloud cover mask, Atmospherically corrected Sentinel 2 and Landsat data
- Availability of the basic datasets as required in the DARD

Moreover, preferred requirements were also expressed which reflects the needs of the HRLC platform and in general of the requirements for the generation of large land cover historical and updated maps.

3 Synthesis of product specification and use cases

3.1 Product and pipeline analysis

At the time of writing of the SRD, the complete pipeline description is not yet the final version in the ATBD. The status of the pipeline will be analysed on the basis of the available official presentations which identifies two main alternatives for the HRLC processing pipeline as:

- **Classification and Decision Fusion** pipeline (also considering deep learning steps) consisting in:
  - Pre-processing steps of Multi-spectral and Radar data
  - Multi-sensor geolocation
  - Classification steps
  - Decision fusion

An example of such pipeline is shown below.

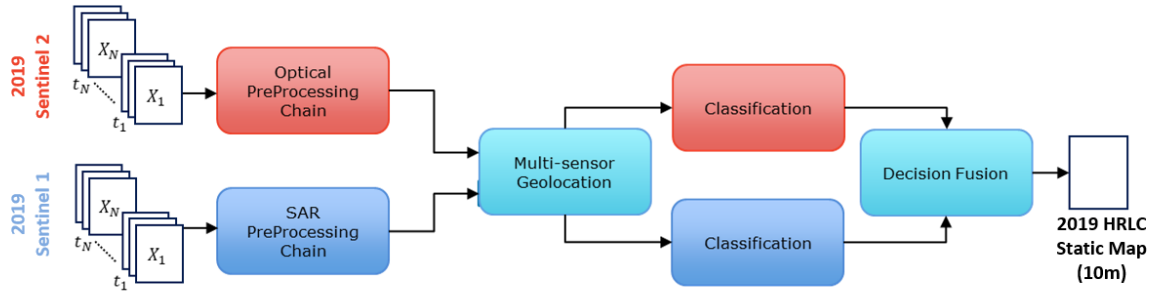

Figure 3: Example Multi-sensor Classification pipeline for static HRLC map generation at 10 m

- **Deep Learning** pipeline consisting in:
  - Pre-processing steps of Multi-spectral and Radar data
  - Single step of Deep Learning classification

An example of such pipeline is shown below.

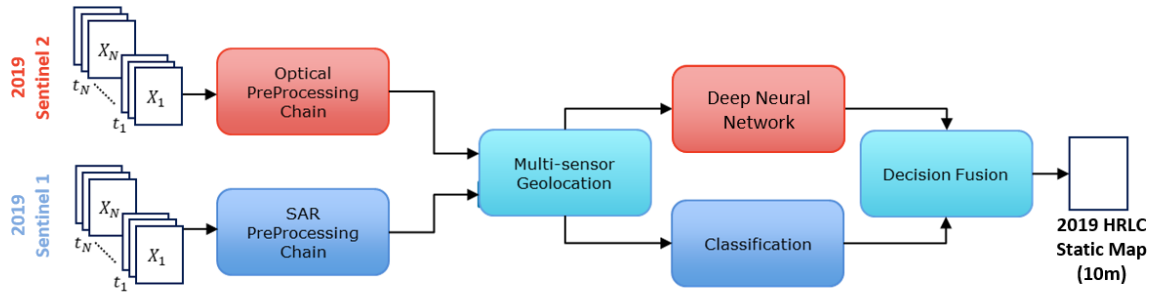

Figure 4: Example Deep Learning pipeline for static HRLC map generation at 10 m

In either case, the pre-processing steps are necessary.

Moreover, it has to be noted that static map generation and historical map generation have similar pipelines even if they have different input data and output (there are differences in resolution and classes identified).

3.2 Use cases and scenarios

As seen in previous paragraph, the HRLC platform shall support both pre-processing steps for Multi-spectral data and Radar data and single/multiple steps of processing ending in the generation of a land cover map. The analysis of use cases and scenarios is performed in order to extract essential requirements for the platform development.

The use cases are organised in pre-processing and processing steps considering also the alternatives of classification described in the previous paragraph.

Being, the final pipelines not yet finalized, the use case analysis has been performed on the basis of the Engineering Team experience in processing and

3.2.1 Pre-processing use cases and scenarios

The PLATFORM shall take care of managing the data pipeline (input/output) for pre-processing of multi-spectral and SAR data. It will be able to execute processing steps like:

- OPTICAL Atmospheric Correction
- OPTICAL Mosaicking
- Spectral filtering
- Spectral adaptation
- Cloud and shadow removal
- **S**AR Co-registration
- SAR Filtering
- SAR MTC (Multi-Temporal Coherence)
- SAR Change Detection

Basic data processing will also provide the capability to store results of the processing pipeline into S3 area allowing on-line processing like band math, drill down over dimensions (time, elevation) and access.

The role of the pre-processing is also to optimize the resources when batch processing jobs are launched. This will be accomplished by providing the possibility to manage several parameters related to the Cloud environment configuration: use of non-guaranteed resources which cost less (e.g. Amazon SPOT instances or Google Preemptible if available on the cloud provider), use of S3-like object storage for the delivery of products and other options related to the resources.

In this sense, two main macro use cases can be identified that cover the above pipeline. The first two steps are meant to be under the “Batch massive pre-processing pipelines for EO data” use case.

Batch massive pre-processing pipelines for EO data			
Use Case ID	UC1		
Date Created	15 June 2019	**Date Last Updated**	15 June 2019
Actors	Main Actors: ● Generic End User of the platform (can be internal or external)		
Main Need	How a user can pre-process in a scalable way EO data for further reuse using existing processing pipelines?		
Description	A user can access the platform using a Web Interface or a programmatic API and add a job for the processing of the data (the job can be directly run for data that are on-line like Sentinel ones or it can be necessary an off-line interaction for other data like those stored in a legacy system). The user can select pre-processing pipelines and datasets among those available below. Once the processing is done, the results are stored into a S3 storage and referenced by a catalogue software which exposes a standard API.		
Specific Details			

V1 Datasets	Sentinel 1	Sentinel 2, Landsat 8
V1 Pipelines	SAR pre-processing generating: orthocorrected images or amplitude/coherence time series	Optical pre-processing generating: atmospheric corrected images, cloud/shadow/water mask
V1 Storage	S3	S3

Table 1: UC1 for batch massive pre-processing pipelines for EO data

3.2.1.1 Pre-processing of multi-spectral optical data

This scenario foresees to do the pre-processing of open data like Landsat 8, Sentinel 2 for large areas in a smaller time-period (1 month). This scenario applies to the HRLC case in the phase of pre-processing prior to the classification for the generation of a land use.

The scenario is to be considered as a simplification and abstraction of the final processing step which will be developed and tuned during the course of the project.

Processing Step	Description	HW Resources	Other ICT notes
Scene selection	The scene selection process facilitates the S2 L1C database query according to the input criteria and yearly baseline check. It transfers the selected scenes into the temporary memory storage ready for further processing.	Virtual machine specs: ● 2 cores ● 16 GB RAM 100 GB storage	High network connectivity (up to 200MB/s) to the database and high load temporary memory storage
Atmospheric correction	The atmospheric correction applies an atmospheric absorption model to the image to produce a Top of Canopy image. This process runs on a single S2 L1C tile and produces S2 L2A tiles.	Virtual machine specs: ● 2 cores ● 16 GB RAM 100 GB storage	

Table 2: Multispectral pre-processing steps

Scenario characteristics (minimum unit resource estimation):

- Minimum AOI extension: single S2 tile (100 sqkm)
- On-shot execution on a full AOI made a single S2 tile for one year

The scenario foresees a single minimum unit of service which is 100 sqkm. This is also the minimum AOI on which the service can be run. The storage is considered temporary, final occupation depends on the desired scale of the product and is minimal with respect to original data.

The pre-processing of this unit can run in 6 hours on a VM node with 2 vCPUs, 16 GB of vRAM and a temporary storage of 100 GB for a stack of 8 images. The final occupation is considerably smaller and is about 20 GB for a full stack of 8 images and about 2 GB for the mosaic.

The pre-processing of Landsat 8 has similar characteristics in terms of resources to be committed.

3.2.1.2 Pre-processing of Sentinel 1

The processing chain, starting from n SAR images (part of the same interferometric stack), will provide a set of SAR ARD data. The processing chain steps, described below, can be executed on Linux Virtual Machine. The following table summarizes the main external libraries required to perform the forest cut and the characteristics of the VM necessary to perform each step.

The scenario is to be considered as a simplification and abstraction of the final processing step which will be developed and tuned during the course of the project.

Processing Step	Description	HW Resources	Other ICT notes
Manager	Given a stack of N images, the Manager will select one as geometric reference and will activate a Virtual Machine (VM) for each of the remaining N-1 in order to perform the image co-registration. It will also identify the set of pairs useful for the coherence map computation and for each of them it will invoke the coherence module.	Virtual machine specs: • 4 cores • 48 GB RAM 100 GB storage	
Coregistration module	This module will download the image pair (master and slave) from local storage and will perform the ingestion of Sentinel-1 data: raster data, annotation auxiliary file and orbital information. The output of this module will be the co-registered slave image (e.g. warped over the geometrical reference grid) that will be uploaded on the storage. This module will be executed on a single VM	Virtual machine specs: • 10 cores • 16 GB RAM • 100 GB storage	High network connectivity (up to 200MB/s) to the database and high load temporary memory storage
Coherence module	This module will download the co-registered image pairs from local storage and will compute the amplitude, of each image, and the interferometric coherence map. The output of this module will be the image amplitudes and coherence map ortho-rectified; these results will be uploaded on the storage. This module will be execute on a single VM.	Virtual machine specs: • 10 cores • 16 GB RAM • 100 GB storage	High network connectivity (up to 200MB/s) to the database and high load temporary memory storage

Table 3: Sentinel 1 processing steps

Scenario characteristics (minimum unit resource estimation):

- Minimum AOI extension: from 3200 to 6400 sqkm wide.
- On-shot execution in correspondence S-1A and S-1B acquisition over the area for one year. The calculation is done on a moving window of 24 S1 scenes.

The scenario foresees a single minimum unit of service which is of about 40 km * 80 km. This is also the minimum AOI on which the service can be run. The storage is considered temporary, final occupation depends on the desired scale of the product and is minimal with respect to original data.

The pre-processing of this unit can run in 4 hours on a VM node with 10 vCPUs, 16 GB of vRAM and a temporary storage of 2,4 TB for a stack of 24 images. The final occupation is considerably smaller and is about 24 GB for a minimum stack of 24 images (minimum unit of service).

3.2.2 Processing use cases and scenarios

The will take care of managing the data pipeline (input/output) for a classification task that, in the present use case analysis is considered only for the Static Map while the generation of the Historical Map is quite similar with a reduced resource need due to the lower resolution of the input/output products.

Classification of stack of multi-sensor images (Classification and Fusion or Deep Learning)			
Use Case ID	UC1		
Date Created	15 June 2019	**Date Last Updated**	15 June 2019
Actors	Main Actors: • Generic End User of the platform		
Main Need	How a user can launch a classification task on a set of ARD (Analysis Ready Datasets) coming from pre-processing.		
Description	A user can access the platform using a Web Interface or a programmatic API and add a job for the processing of the data. The user can select processing pipelines and datasets among those available. Once the processing is done, the results are stored into a S3 storage and then published in the distribution system. The processing steps are creating a single output from multiple input which consists in both multi-sensor and time-series data. The output is the CCI HRLC static map as described in the PSD.		
Specific Details			
V1 Datasets	ARD for Sentinel 1	ARD Sentinel 2	
V1 Pipelines	The pipeline can be one of the two described above, i.e.: • Classification and Fusion pipeline • Deep Learning pipeline Depending on the specific chain (SAR, Multispectral). Then the different chains will be fused.		
V1 Storage	S3		

Table 4: UC2 Classification of stack of multi-sensor images

To this extent, the following procedure illustrates the processing steps and functionalities that the PLATFORM shall be able to support. In addition, the paragraph provides some highlights of certain practical scenarios that the PLATFORM shall manage.

3.2.2.1 Classification and fusion processing

The following processing chain description makes hypothesis on the processing steps starting from the paragraph 3.1. It also tries to make assumptions on the HW resources needed for each step. This will allow to

The scenario is to be considered as a simplification and abstraction of the final processing step which will be developed and tuned during the course of the project.

Processing Step	Description	HW Resources	Other ICT notes
Multisensor Geolocation	This step will generate coregistered geolocated from multisensor ARD (Sentinel 1, Sentinel 2, Landsat 8 etc.) using modern techniques. This module will be execute on a single VM within 0,5 hour.	Virtual machine specs: ● 2 cores ● 32 GB RAM 100 GB storage	
HRLC Optical data classification	This module will take as input the sentinel 2 (atmospherically corrected and cloud masked), plus the training samples generated in the previous step. Obtained HRLC layer will be uploaded on the repository storage. This module will be execute on a single VM within 0,5 hour.	Virtual machine specs: ● 4 cores ● 48 GB RAM ● 100 GB storage	High network connectivity (up to 200MB/s) to the DIAS database and high load temporary memory storage
HRLC SAR data classification	This module will take as input the sentinel 1 (amplitude and coherence data), plus the training samples generated in the previous step. Obtained HRLC layer will be uploaded on the repository storage. This module will be execute on a single VM within 1 hour.	Virtual machine specs: ● 4 cores ● 64 GB RAM 100 GB storage	The amount of memory depends on the number of images in the time-series. Generally, SAR time series are denser that the optical one due to the fact that there is no Cloud coverage.
Decision Fusion	This module will implement a decision fusion step using different techniques Class-based rules like Consensus Theory (linear/ log opinion pool, etc.) or Markov/conditional random fields (MRF / CRF).	Virtual machine specs: ● 2 cores ● 16 GB RAM 100 GB storage	

Table 5: Classification and fusion processing steps for the classical methodology

Scenario characteristics (minimum unit resource estimation):

- Minimum AOI extension: single S2 tile (100 sqkm)
- Available ARD of data over the area

The scenario foresees a single minimum unit of service which is 100 sqkm. This is also the minimum AOI on which the service can be run. The storage is considered temporary, final occupation depends on the desired scale of the product and is minimal with respect to original data.

The processing in this case involve different resources with different size and the best configuration will be defined once the processing chain is fully defined. The estimation made in the table are reasonable example of HW resources needed for each step.

3.2.2.2 Deep Learning processing

Proc. Step	DESCRIPTION	HW RES	OTHER ICT
HRLC classification	This module will take as input the sentinel 1 (amplitude and coherence data), sentinel 2 (atmospherically corrected and cloud masked) and landsat 8 (atmospherically corrected and cloud masked), plus the training samples generated in the previous step. Obtained HRLC layer will be uploaded on the repository storage. This module will be execute on a single VM within 2 to 6 hour (depending on the model performance which can varies over different areas).	Virtual machine specs: • 24 cores • 128 GB RAM • 100 GB storage • 2x16GB GPU	High network connectivity (up to 200MB/s) to the DIAS database and high load temporary memory storage. Need of additional GPU can be estimated only after the fine tuning of the Deep Learning model as the number of layers strongly influence this number.

Table 6: Classification processing steps for the Deep Learning scenario

Scenario characteristics (minimum unit resource estimation):

- Minimum AOI extension: single S2 tile (100 sqkm)
- Available ARD of data over the area

The scenario foresees a single minimum unit of service which is 100 sqkm. This is also the minimum AOI on which the service can be run. The storage is considered temporary, final occupation depends on the desired scale of the product and is minimal with respect to original data.

The processing in this case involve different resources with different size and the best configuration will be defined once the processing chain is fully defined. The estimation made in the table are reasonable example of HW resources needed for each step.

4 Detailed system requirements

4.1 Requirements on the generated output products

REQ-01 - The PLATFORM shall be able to generate The format of the product will follow the guidelines of the CCI Data Standards v2.0 (and the upcoming v2.1 when finalised). The delivery formats are the two data format encodings commonly used in the Climate Change modelling community (NetCDF) and Land Cover community (GeoTIFF). The standard includes both indication on the data content specifically for NetCDF encoding and indications on naming and on metadata.

Considering the size of the final products which is estimated to be several GB, the PLATFORM shall provide the capability to split the final products in tiles following the MGRS Military Grid Reference System tiling (or a similar Grid Reference System) which is a well-known tiling system already used in Sentinel 2 delivery.

REQ-02 -The PLATFORM shall allow to hosts metadata information as INSPIRE compliant metadata records to be provided to the Data Portal team to be included in the portal services.

REQ-03 -The PLATFORM shall at least support the projection of the products according to:

- Unit: degree (supplier to define representation)
- Geodetic CRS: WGS 84
- Datum: World Geodetic System 1984
- Ellipsoid: WGS 84
- Prime meridian: Greenwich
- Data source: OGP
- Information source: EPSG. See 3D CRS for original information source.
- Revision date: 2007-08-27
- Scope: Horizontal component of 3D system. Used by the GPS satellite navigation system and for NATO military geodetic surveying.
- Area of use: World.
- Coordinate system: Ellipsoidal 2D CS. Axes: latitude, longitude. Orientations: north, east. UoM: degree

REQ-04 -The PLATFORM shall support transformation and conversion tools to allow the above mentioned specification and in general shall support future changes with flexibility in the transformation and conversion of geodata.

REQ-05 -The PLATFORM shall support the Encoding formats specified in 3.1.3 and 3.2.3 and follow the CCI Data Standards v2.0 in which the structure of the data is specified in terms of variables (bulk data) and attributes (ancillary data).

4.2 Requirements on input data

4.2.1 High resolution multispectral data

REQ-06 -The PLATFORM shall provide access to multi-source satellite/aerial EO data including, as a minimum, the following multispectral datasets from the suggested sources/endpoints. It is recommended that access to the same resource form multiple endpoints is implemented in order to increase the overall system redundancy.

Data volume that the PLATFORM is required to handle

Dataset	Product	Volume requirement for Static HRLC	Volume requirements for Historic HRLC	Total Volume requirements (TB)
Multispectral optical	Sentinel-2	Amazon 65 TB Africa 111 TB Siberia 158 TB		334
	Landsat-5		Amazon 5 TB Africa 3 TB Siberia 2 TB	10
	Landsat-7		Amazon 8 TB Africa 6 TB	17.5

			Siberia 3.5 TB	
	Landsat-8		Amazon 6 TB Africa 6.3 TB Siberia 5 TB	17.3
SAR				6.5
	Sentinel-1 L1 IW GRD HR Product	6.5 TB	4.7 TB	4.7

Static High Resolution Land Cover Requirements

Dataset	Processing level	Source(s)
Sentinel-2	L1C (optional) L2A (mandatory)	DIAS, ESA SciHub, AWS, Google Cloud
Landsat-8	L2A (mandatory)	AWS, Google Cloud
Landsat legacy	L2A (mandatory)	Google Cloud, AWS (if available)

Historical High Resolution Land Cover Requirements

Dataset	Processing level	Source(s)
Landsat-5	L1TP (Level-1 Precision Terrain) Tier 1 and L2SP	USGS
Landsat-7	L1TP (Level-1 Precision Terrain) Tier 1 and L2SP	USGS
Landsat-8	L1TP (Level-1 Precision Terrain) Tier 1 and L2SP	USGS

4.2.2 High resolution SAR data

REQ-07 - The PLATFORM shall provide access to multi-source free and open satellite/aerial EO data including, as a minimum, the following multispectral datasets from the suggested sources/endpoints. It is recommended that access to the same resource form multiple endpoints is implemented in order to increase the overall system redundancy.

Static High Resolution Land Cover Requirements

Dataset	Processing level	Source(s)
Sentinel-1	L1B GRD (mandatory)	DIAS, ESA SciHub, Alaska Satellite Facility (ASF) and, if available, AWS and Google Cloud

4.2.3 Validation data requirements

REQ-08- The PLATFORM shall be able to ingest the following data sources:

- WorldView-1/-2/-3

- Pleiades
- GeoEye-1
- IKONOS
- SPOT (1 to 4)

REQ-09 - The PLATFORM shall be able publish validation data in the distribution system.

4.3 Requirements on ancillary input data

Ancillary data are used by the processing chains for pre-processing (Sentinel 1 orbits files, DEM) and processing (land cover used for testing the classification).

The pre-processing tasks will require some ancillary data in terms of non-geodata and geodata. Some of the ancillary data shall be made available in the data repository, others can be retrieved on-request.

The processing tasks can require the use of land cover data for validation purposes and for training purposes.

4.3.1 Ancillary data useful for pre-processing

REQ-10 - The PLATFORM shall make available on-request the Sentinel 1 orbits files for any scene/product

The PLATFORM shall make available on-request the SRTM DEM 4.1

4.3.2 Ancillary data useful for processing

REQ-11 - The PLATFORM shall make available on-request the Open Street Map data

REQ-12 - The PLATFORM shall make available in the repository the following Land Cover data sources

- Global 100m Land Cover, https://lcviewer.vito.be/download
- Global Urban Footprint, DLR
- Global Human Settlement Layer (GHS BUILT-UP GRID S1), JRC
- Global Surface Water, JRC

REQ-13 - The PLATFORM shall make available in the repository additional ancillary data sources, as they are required. The addition of a new ancillary data shall require the execution of a procedure or an automated script by the PLATFORM administrator.

4.4 Requirements on metadata

Metadata are necessary for the data selection and filtering once the processors are ready to be used on the area of interests. The metadata elements to be considered are different for optical missions and radar missions.

REQ-14 - The PLATFORM shall provide access to the metadata of the whole datasets described in paragraph 4.1 (HRLC output products), paragraph 4.2 (input satellite data and validation data) and paragraph 4.3 (ancillary data).

REQ-15 - The PLATFORM shall provide capabilities to handle and store metadata on the datasets.

REQ-16 - The PLATFORM shall be able to provide metadata for processing chains and processors.

4.5 Requirements on pre-processing platform

The processing platform is designed to apply in an optimized way the pipelines developed and tested by the Science Team. As a common approach agreed with the Science Team, the platform will not be operated during the development phase. This because the generation of the HRLC products is a big-data problem that must be migrated to modern Cloud environment such as DIAS. For this reason and to keep the IaaS budget reduced, the development and selection of the best pipeline is done by the Science team during a prolonged development phase.

Nevertheless, the architecture that is highlighted is meant to allow in principle the continuous improvement of the pipelines and this is reflected in the requirements and in the capabilities that the PLATFORM will provide. This because the future use of the platform will be to also provide the means to execute new algorithms, new versions of algorithms or new parameterisation, for testing, intercomparison and evaluation purposes.

4.5.1 Data pre-processing and processing facility

REQ-17 - The PLATFORM shall allow Users to create processing nodes as task.

REQ-18 - The PLATFORM shall allow Users to execute the created task.

REQ-19 - The PLATFORM shall allow to configure new pre-processing chains as Dockerfiles

REQ-20 - The PLATFORM shall be able to create Docker images. This allows the integration of algorithms (also already existing) and tools without the need to re-develop them in a common programming languages.

REQ-21 - The PLATFORM shall be able to instantiate Docker container with optimized resource

REQ-22 - The PLATFORM shall be able to optimize the IaaS resource consumption for the different processors. This will be done by selecting the best instance for the specific processors (or by configuring the Docker) able to use the resources at the higher level (high percentage of consumption of CPU and Memory during the processing). The information will be provided the Science Team and registered in the processors metadata.

REQ-23 - The PLATFORM shall be able to tune the IaaS resource configuration for each processor.

REQ-24 - The PLATFORM shall be able to launch predefined commands with parameters specified with an API interface in the Docker container

REQ-25 - The PLATFORM shall be able to collect outputs and logs from the Docker container instance

REQ-26 - The PLATFORM shall be able to manage failures in the processing and notify the user

REQ-27 - The PLATFORM shall provide a full automated DevOps process to manage the update of the processors and management of the versions of the processors (together with an integrated GIT repository)

4.5.2 Data pre-processing (SAR)

REQ-28 - The PLATFORM shall allow Users to exploit SAR data pre-processing pipelines based pre-processing chains developed by the Scientific Team and Engineering Team. The SAR Data pre-processing pipeline shall be composed at least by the following steps (after data ingestion):

- Classes morphological erosion
- Training set extraction
- Despeckling filter (Lee)
- Multi-temporal despeckling filter

REQ-29 - The PLATFORM SAR data pre-processing shall exploit tools such as:

- Open SW (e.g. Sentinel toolbox)
- Other solutions developed/integrated by the Scientific Team and Engineering Team

REQ-30 - The PLATFORM SAR data-pre-processing shall be available either as an end-to-end workflow (with configurable options to select the preferred engine or the steps to be performed or to pass expert parameters to the pipeline) or as single building blocks for further chaining in custom models.

REQ-31 - In either cases, the PLATFORM SAR pre-processing options will be configurable and will be implemented taking into account the level of parametrization of the used tools (e.g. Sentinel toolbox, other tools).

REQ-32 - The PLATFORM SAR data pre-processing shall be able to access and use any auxiliary data needed for the pre-processing step execution.

REQ-33 - The PLATFORM SAR data pre-processing shall be able to process data described in this document (paragraph 4.2.2).

REQ-34 - The PLATFORM SAR data pre-processing pipeline shall allow to generate ARD ready for being ingested by the processing pipelines (HR Land Cover classification pipelines).

4.5.3 Data pre-processing (Multi-spectral)

REQ-35- The PLATFORM shall be able to ingest Sentinel 2 L2A data as input to the processing chain as ESA already provides BOA data for outside Europe for 2019.

REQ-36 - The PLATFORM shall be able to ingest Landsat 5/7/8 L2SP Surface Reflectance products that include also Cloud Mask calculated with USGS provided FMask algorithm.

REQ-37 - The PLATFORM shall provide capability to further chain above products in custom models. These building blocks shall also be usable in batch configurations to process large amounts of data in advance of the classification process.

There shall be two main options to approach the BOA calculation process:

1. Process the full archive of required data in advance and make the BOA data available on the processing infrastructure storage for the classification process.
2. Include the BOA calculation as a step in the classification process and discard the BOA data after is has been processed.

4.5.3.1 Sen2Cor

Sen2Cor is a processor for Sentinel-2 Level 2A product generation and formatting. It performs the atmospheric, terrain and cirrus correction of Top-Of- Atmosphere Level 1C input data. Sen2Cor creates Bottom-Of-Atmosphere, optionally terrain- and cirrus corrected reflectance images. Additionally, Aerosol Optical Thickness, Water Vapor, Scene Classification Maps and Quality Indicators for cloud and snow probabilities. Its output product format is equivalent to the Level 1C User Product: JPEG 2000 images, three different resolutions, 60, 20 and 10 m.

Sen2Cor installation packages are provided as standalone installers. Two different versions of Sen2Cor are available at the time of writing:

- Sen2Cor_v2.8 is the latest release supporting Sentinel-2 L1C data generated with the current (14.5) and previous (14.2) Products Specification Document, as input. Sen2Cor_v2.8 release contains new features and improvements respect to Sen2Cor_v2.5.5.
- Sen2Cor_v2.5.5 is the previous release and it is needed if the user intends to process old Sentinel-2 L1C data generated with the Products Specification Document older than 14.2 and not reprocessed by ESA.

The Sentinel 2 data for the AOIs over one year amount to approximately **550k products** which require **roughly 350-400 TB of storage**. The processing of L2A data for one product requires in average between 15 to 30 minutes on a 2CPU and 8GB virtual machine which sums up to a processing time of **up to 250k hours**.

To reduce this significant storage and processing requirement, a filtering of unsuitable products can be considered. Such unsuitable products may be ones with quality issues, high cloud coverage or limited data coverage within the tile footprint.

4.5.3.2 Lansat surface reflectance project

The Landsat 5 & Landsat 7 data shall amount to approximately **125k scenes** and require **roughly 20 TB** of storage capacity. The Landsat 8 data shall amount to approximately **30k scenes** and require **roughly 30 TB** of storage capacity.

At the time of writing, the information about Landsat availability on the DIAS showed that other than limited Landsat 8 data, the majority of required data was not available in the archives, which makes procuring the data from USGS necessary. The process of retrieving data from USGS includes an ordering step in which the processing level is to be specified. Thus, it could be an option to order the BOA data directly, which would render the need for processing BOA data unnecessary.

If processing of BOA data is required there are two different tools available:

LEDAPS

Landsat 4-5 TM and Landsat 7 ETM+ Surface Reflectance are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm, a specialized software originally developed by NASA Goddard Space Flight Center (GSFC) and the University of Maryland (Masek et al., 2006).

The software applies Moderate Resolution Imaging Spectroradiometer (MODIS) atmospheric correction routines to Level-1 data products. Water vapor, ozone, geopotential height, aerosol optical thickness, and digital elevation are input with Landsat data to the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) radiative transfer models to generate top of atmosphere (TOA) reflectance, surface reflectance, TOA brightness temperature, and masks for clouds, cloud shadows, adjacent clouds, land, and water.

LaSRC

Landsat 8 SR data products are generated from specialized software called Landsat 8 Surface Reflectance Code (LaSRC). The original Landsat 8 LaSRC algorithm was developed by Dr. Eric Vermote, National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and was modified by staff at USGS Earth Resources Observation and Science (EROS) center.

LaSRC generates Top of Atmosphere (TOA) Reflectance and TOA Brightness Temperature (BT) using the calibration parameters from the metadata. Atmospheric correction routines are then applied to Landsat 8 TOA Reflectance data, using auxiliary input data such as water vapor, ozone, and Aerosol Optical Thickness (AOT) retrieved fromModerate Resolution Imaging Spectroradiometer (MODIS), and digital elevation derived from GTOPO5 to generate Surface Reflectance. The result is delivered as the Landsat Surface Reflectance (BOA) data product.

4.5.4 Data access and input/output

REQ-38 - The PLATFORM shall allow Users to handle multi-mission EO data and ancillary data from the PLATFORM data & metadata through a single search interface based on geographic and non-geographic filters.

REQ-39 - The PLATFORM User shall be able to refine the search results, group them by mission and to sort them based on available metadata (e.g. data, cloud coverage).

REQ-40 - The PLATFORM shall take as input (mainly from the data repository):

- Most common raster and vector formants (see GDAL/OGR supported formats for reference, including OGC web services)
- Data from satellite EO missions mentioned in this document

REQ-41 - The PLATFORM shall be able to read/import:

- Data from local file systems or from storage areas which are connected through the internet (e.g. FTP, SFTP)
- Cloud storage (e.g. AWS s3 storage or Google Cloud equivalent)

In general, it is of outmost importance that the PLATFORM is capable to maximize read/write throughput through the use of distributed file systems options (also made available by cloud providers, e.g. chunks, parallel access to the object storage).

REQ-42 - The PLATFORM shall be able to write output as:

- Data from local file systems or from storage areas which are connected through the internet (e.g. FTP, SFTP)
- Cloud storage (e.g. AWS s3 storage or Google Cloud equivalent)

REQ-43 - The PLATFORM shall be able to manage long term archiving by the use of different layer of S3 storage (commonly referred as Cold storage which is less expensive to Hot storage which is more expensive).

REQ-44 - The PLATFORM will provide a local storage backup repository for the output products.

4.6 Requirements on processing environment

Based on the processing requirements, the most suitable processing environment shall be selected. To identify the most suitable one, the different DIAS shall be evaluated against the following criteria. The DIAS provides typically an enhanced IaaS environment with attached services like standard OGC CSW/OpenSearch catalogue and OGC WMS; moreover, they provide access to Copernicus data (Sentinel 1/2/3) and other missions like Landsat.

For this reason, the requirements are divided in two categories: IaaS and Data. A minimal/preferred set of requirements is declared.

REQ-45 - The PLATFORM must be hosted in the most suitable IaaS environment with the maximum number of environment requirements satisfied. Environment requirements are described in the following sub-paragraphs.

4.6.1 Minimal IaaS requirements

The environment shall provide an EC2 like API (https://github.com/openstack/ec2-api, https://docs.aws.amazon.com/AWSEC2/latest/APIReference/), possibly working against the standard library boto3 (https://github.com/boto/boto3). The API shall allow to manage instances (creation, deletion, Spin up/down …) using the boto3 library which, despite being developed for Amazon AWS, it is widely considered a de facto standard for IaaS interaction using Python programming language.

The environment shall provide scalability to allow processing of multiple (at least hundreds) of scenes in parallel.

The environment shall provide an object storage with an S3 API with up to tens of TB storage capacities. The API shall allow to manage objects (creation, deletion, download …) using the boto3 library which, despite being developed for Amazon AWS, it is widely considered a de facto standard for IaaS interaction using Python programming language.

The environment shall provide instances with GPU processing units in order to allow the training of deep learning models.

The environment shall provide also big instances for CPU processing (> 128 GB Memory, > 48 vCPU) in order to allow simple processing of classification tasks involving multiple images (e.g. time series).

The environment shall provide at least a 99.999 SLA and support

The environment shall provide adequate system and API documentation

The environment shall provide per-hour billing at minimum

4.6.2 Minimal Data requirements

The environment shall provide the full Sentinel 1/2 archive for 2019

The environment shall provide access to a OGC CSW/OpenSearch catalogue

The environment shall provide the capability to retrieve additional public data

4.6.3 Preferred IaaS requirements

The environment shall preferably provide Virtual machine images with pre-installed relevant software related to Remote Sensing. Among the others:

- ESA SNAP toolbox for Sentinels and historical missions
- Conda Python environment
- Keras/Tensorflow with already configured GPU NVIDIA drivers

The environment shall preferably provide flexible instance types (different RAM/CPU settings for either computational or RAM heavy processes)

The environment shall preferably provide processing/computation optimized configuration instances. These instances have commonly a parity between CPU and Memory (e.g. 1 CPU paired with 1 GB)

The environment shall preferably provide memory optimized configuration instances. These instances have commonly 1/4 ration between CPU and Memory (e.g. 1 CPU paired with 4 GB)

The environment shall preferably provide a container service

The environment shall preferably provide Spot instances (In order to reduce processing costs)

The environment shall preferably provide per-minute/per-second billing. This will facilitate the parallelism of pre-processing as there is the possibility to launch full parallel processing (e.g. of the duration of less than 1 hour) without being billed the full hour.

4.6.4 Preferred Data requirements

The environment shall preferably provide data in COG (Cloud Optimized GeoTIFF)

The environment shall preferably provide precalculated Cloud Mask in the case of Multi-spectral data

The environment shall preferably provide Landsat historical archives since Landsat 4 up to Landsat 8

The environment shall preferably provide support/courtesy/functionalities when it comes to retrieve and host Landsat data which is not available in the DIAS archive

The environment shall preferably provide Sentinel 2 L2A archive for 2019

The environment shall preferably provide support/courtesy/functionalities when it comes to processing and hosting of required L2A data which are not provided by ESA

The environment shall preferably provide ancillary data archives as mentioned in the paragraph 4.3

4.7 Requirements on distribution services

The distribution environment will have the responsibility to deliver the product to the CCI Climate Open Data portal and as services (SFTP, OGC OWS etc.). User profiles shall be also implemented in order to allow different profiles to perform operations like to evaluate the quality of the pre-processed data by means of indicators to be designed and assessed for each data type.

4.7.1 Delivery and distribution services

REQ-46 - The PLATFORM shall allow the Users to download the product from a SFTP service at minimum.

REQ-47 - The PLATFORM shall allow the Users to search on the catalogue. In general, the metadata defined in paragraph 4.4 will be hosted in the catalogue.

REQ-48 - The PLATFORM shall allow Users to quickly display data products in a GIS like environment through OGC WMS services.

REQ-49 - The PLATFORM shall provide access through a Web interface to visualise data products through OGC WMS services.

REQ-50 - The PLATFORM shall provide the following capabilities through the Web interface:

- Possibility to configure and save custom maps and mashups
- Possibility to add small shapefiles through upload
- Possibility to add external WMS to the map
- Possibility to exploit the temporal dimension of data through a time bar

REQ-51 - The PLATFORM shall be able to interact with the CCI Climate Open Data portal.

4.7.2 User management services

REQ-52 - The PLATFORM shall implement a user profiling system able to manage at least the following profiles:

- Generic User: a user profile which can access all the PLATFORM datasets and functionalities for the visualisation of maps and mashups and of OGC OWS services through GIS software.
- Power User: a user profile which can access all the PLATFORM datasets apps and functionalities, with the configure new maps and mashups. The PLATFORM Power User has the possibility of creating new datasets and new workflows using a development API.
- Administrator: full control of the PLATFORM, with the possibility of creating new users and managing users permissions.

4.8 General requirements

4.8.1 Performance requirements

REQ-53 - The PLATFORM shall be able to manage up to hundreds of processing nodes for the pre-processing phase.

REQ-54 - The PLATFORM shall be able to generate the required CCI HRLC products in a reasonable time with maximum elapsed of 1 month for each AOI.

REQ-55 - The PLATFORM shall be able to sustain the delivery of the products through the Web interface to up to 20 users.

4.8.2 Monitoring requirements

REQ-56 - The PLATFORM used to run the CCI HRLC product generation shall be able to monitor the IaaS resources consumed, with the possibility of setting alerts in case certain thresholds are overpassed.

REQ-57 - The PLATFORM used to run the CCI HRLC product generation shall be able to provide detailed reports about the resources consumed by each run in order to provide all the necessary elements for the analysis and forecast of money expenditure.

REQ-58 - The PLATFORM used to run the CCI HRLC product generation shall be monitored in order to ensure that all the necessary services are properly running. In case of incidents or malfunctioning of internal pipeline or of the external services (Catalogue, Download services), an alert shall be raised to e-GEOS Administrators.

4.8.3 Standards Requirements

REQ-59 - The following standards have to be take into consideration for the development of the PLATFORM.

Reference	Title
[STD-1]	OGC 05-007r7, Web Processing Service, OGC Implementation standard, 2007-06-08
[STD-2]	OGC 07-063r1, Web Map Services – Profile for EO Products, Best practice, Version 0.3.3, 05/11/2009.
[STD-3]	OGC 10-032r8, OGC OpenSearch Geo and Time Extensions, http://www.opengeospatial.org/standards/opensearchgeo
[STD-4]	OGC 10-140, OGC WCS 2.0 IS - Earth Observation Application Profile, version 1.0.0, 14/02/2014.
[STD-5]	OGC 10-157r4, Earth Observation Metadata profile of Observations & Measurements, Version 1.1, 09/06/2016, http://docs.opengeospatial.org/is/10-157r4/10-157r4.html.
[STD-6]	OGC 12-084r2, OGC OWS Context Atom Encoding Standard, https://portal.opengeospatial.org/files/?artifact_id=55183
[STD-7]	OGC 13-026r8, OGC OpenSearch Extension for Earth Observation Products, http://docs.opengeospatial.org/is/13-026r8/13-026r8.html.
[STD-8]	OGC 13-043, Download Service for Earth Observation Products, Version 1.0, 31/01/2014.
[STD-9]	OGC 14-055r2, OGC OWS Context GeoJSON Encoding, 30/05/2016, http://docs.opengeospatial.org/is/14-055r2/14-055r2.html
[STD-10]	OGC 14-065, Web Processing Service, Version 2.0.1, 25/03/2016.
[STD-11]	OGC 17-003, OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard, Version 0.0.8, 06/09/2017.
[STD-12]	OGC 17-047, OGC EO OpenSearch Response GeoJSON(-LD) Encoding Standard.
[STD-13]	OGC Network Common Data Form (NetCDF) Core Encoding Standard version 1.0

5 Requirements traceability

Current requirement traceability is performed vs the Tender Requirement identified in the Tender specification:

Tender Requirement	Description	Traceability
TR-12	The prototype system to have the capacity to handle data from multiple sensors at different spatial resolutions with the capacity to deal with large high resolution data volumes (e.g. multi-temporal S1, S2, Landsat 8) and very high resolution data	Par. 4.2

	volumes (e.g. Pleaides, SPOT-6). Potential contributing sensors are listed in Section 9.	
TR-13	The HRLC-System shall provide the scientists a configurable, flexible, agile and open HRLC CCI workflow for managing the evolution of the HRLC ECV. It shall have the following capabilities: ● Multi-sensor full mission (re-)processing for all HRLC ECV products ● Systematic, data-driven processing, allowing rapid ingestion of new data ● Configuration management (processor versions, auxiliary data, input and output data, etc.) ● Capability to integrate (existing) tools developed in different programming languages ● Capability for algorithm developers to easily trial plug in, without recompilation where practical, and execute new algorithms, new versions of algorithms or new parameterisation, for testing, intercomparison and evaluation purposes ● Support efficiently all the needs of the scientific algorithms and the new development	Par. 4.5
TR-15	The HRLC-System shall include data access, ingestion, product conversion tools and distribution functionality, and shall address long-term archiving of both input and output products. Both requirements shall meet the generic needs outlined in the main body of the CCI SoW.	Par 4.5.4
TR-16	The design of the HRLC-system shall be based on experience from previous CCI projects (e.g. LandCover_cci and Fire_cci) and from other research and operational processing schemes (e.g. Copernicus).	Par. 4.6 as Third Party service of DIAS in the Copernicus scheme
TR-17	The HRLC-System shall have data access interfaces to handle efficiently, and in a standardised fashion, the data streams from Sentinel 1 and 2. The interfaces shall support the possibility of cross-ECV synergies.	Par. 4.6.2 as a requirement of the DIAS environment where the system will be deployed
TR-18	The contractor shall link with the CCI portal and data analysis/visualisation tools available through the CCI Toolbox	Par. 4.7