World AgroCommodities (WAC)

WorldAgroCommodities

D1.2 Requirements Baseline

ESA Ref:	4000131862/20/I-DT
Del. No.:	ESA_WAC_D1.2_Requirement Baseline
Issue/Rev.:	1.1
Date:	31.01.2025

Lead:

GAFAG an e-GEOS (ASI / Telespazio) Company

Partners:

vito TerraSphere WAGENINGEN UNIVERSITY & RESEARCH MERIDIA GFZ Helmholtz Centre Potsdam e-geos AN ASI/TELESPAZIO COMPANY

Financed by:

esa

Consortium Partners

No.	Name	Short Name	Country
1	GAF AG	GAF	Germany
2	Vito	VITO	Belgium
3	Wageningen University	WU	The Netherlands
4	GFZ – Helmholtz Centre Potsdam	GFZ	Germany
5	TerraSphere	TerraSphere	The Netherlands
6	Meridia	Meridia	The Netherlands
7	e-geos	e-geos	Italy

Disclaimer:

All texts, photos and graphic designs used in this document are protected by copyright. Should you wish to use parts of them, please contact GAF AG. GAF AG will then, if necessary, establish contact with the author or authorised user.

Summary

This document presents the findings for the D1.2 Requirements Baseline.

	Affiliation/Function	Name	Date
Prepared	GAF AG, GFZ	C. Sannier, S.Gomez, Katja Berger, Martin Herold	16/12/24
Contributions	VITO, WU	Hans Vanrompay, Jeroen Degerickx	20/12/24
Reviewed	GAF AG	S. Gomez	02/01/25
Approved	GAF AG, Project Coordinator	C. Sannier	02/01/25

The document is accepted under the assumption that all verification activities were carried out correctly and any discrepancies are documented properly.

Distribution

Affiliation	Name	Copies
ESA	Z. Szantoi	electronic copy
Partners	Consortium Members	electronic copy

Document Status Sheet

Issue	Date	Details
1.0	02/01/25	First Document Issue
1.1	31/01/25	Second Document issue

Document Change Record

#	Date	Request	Location	Details
1	31/01/25	Revision	Whole document	Minor corrections
2				

Executive Summary

This report provides a summary of the activities undertaken in the first quarter of the project which had an aim to understand the user requirements and identify the test and demonstration sites. Additionally there has been an examination of the Earth Observation (EO) tools/approaches available for meeting the EU Deforestation Free Supply Chain Regulation (EUDR) and this in combination with the user requirements provided the basis for presenting a technical approach the project will take. Thus the report presents a framework for the Phase 1 of the project activities. The main objectives of the report are therefore to:

- provide a comprehensive review and analysis of the underlying policy framework addressed by
- the project;
- present the target user group and of the user needs and challenges to be addressed by the project, based on the guided interviews with the Competent National Authorities (CNAs); this includes a risk assessment of the user requirements in the context of the project.
- A review of the current state-of-the-art literature/studies on relevant EO algorithms, methods, models addressing the project needs, and the major knowledge gaps and challenges.
- A proposed technical approach to achieve the EO-integrated solution to be developed and validated; and the specifications of the test and demonstration sites that can form the basis for the Use Case studies to be conducted to demonstrate the adequacy, robustness, scalability and usefulness of the proposed solution.

As part of this requirement baseline assessment a policy review focussing on the EUDR was conducted followed by a summary of the User Requirements emanating from the extensive consultation with CNAs through guided interviews a first living lab and an analysis of the potential risks associated with some of the outlined requirements. The latest developments were reviewed related to relevant technical solutions currently available to meet the different user requirements, as well as the gaps in the current methodologies which then led to a description of a proposed technical approach for the current project. This was then followed with a structured approach on how the test and demonstration sites were selected for the different commodities.

The key gaps identified as part of the review of the state-of-the-art are primarily linked to the following main topic:

- Single crop focus as opposed to tackling multiple commodities
- Link to deforestation not always assessed or explicit
- Limited focus on commodity expansion over several years
- Smallholder production systems (e.g. Agroforestry) are not always properly characterised as they cover small areas, are difficult to map and can often be confused with forest areas

Some key methodological challenges were also outlined:

- Data accessibility and pre-processing to ensure analysis ready data with constant quality are available.
- State-of-art ML/DL models require large amount of representative datasets to be applicable globally and their integration in OpenEO workflows also represents a substantial challenge
- While the envisaged Open architecture of the WAC prototype system can integrate private data into its workflows, special attention must be given to addressing the privacy requirements of the CNAs.

The detailed feedback received from CNAs made it possible to outlined the requirements for the WAC protype system following a two-step approach:

1. A rapid potential non-compliance detection system followed by
2. An evidence-based verification system

The requirements expressed were translated in a set of functional and technical characteristics that will be tested in the next step focusing on the benchmarking of methods with the aim to assess the feasibility and limitations of addressing some of these requirements. The final specifications for the WAC prototype system will be documented in the Algorithm Theoretical Baseline Document (ATBD) and

Technical Specifications (TS) and will be implemented over the selected demonstration sites ensuring CNAs and other stakeholders are consulted at each key step during the process.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ATBD	Algorithm Technical Baseline Document
AWS	Amazone Web Service
BLE	Bundesanstalt für Landwirtschaft und Ernährung (Germany
CDSE	Copernicus Data Space Ecosystem
CNAs	Competent National Authorities
CNN	Convolutional Neural Network
DDS	Due Diligence Submission
EC	European Community
EU	European Union
EUDR	European Union Deforestation Regulation
EO	Earth Observation
ESA	European Space Agency
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	Food and Agriculture Organization
FPS Health	Federal Public Service Health, Food Chain Safety and Environment (Belgium)
GAF	GAF AG, Geospatial Service Provider, Germany
GFC	Global Forest Cover
GIS	Geographic Information System
GFZ	GFZ Helmholtz Centre for Geosciences
JRC	Joint Research Centre
ML	Machine Learning
NICFI	Norway's International Climate and Forest Initiative
NVWA	Netherlands Food and Consumer Product Safety Authority
RF	Random Forest
SEF	Stakeholder Engagement Facility
SoW	Statement of Work
TCD	Tree Cover Density
TS	Technical Specifications
UDF	User Defined Function
UHUL	Ústav pro hospodářskou úpravu lesů Brandýs nad Labem (Czech Republic)
VHR	Very high resolution

This page is intentionally left blank!

1 Introduction and Objectives

This report provides a summary of the activities undertaken in the first quarter of the project which had an aim to understand the user requirements and identify the test and demonstration sites. Additionally there has been an examination of the Earth Observation (EO) tools/approaches available for meeting the EU Deforestation Free Supply Chain Regulation (EUDR) and this in combination with the user requirements provided the basis for presenting a technical approach the project will take. Thus the report presents a framework for the Phase 1 of the project activities. The main objectives of the report are therefore to:

- provide a comprehensive review and analysis of the underlying policy framework addressed by
- the project;
- present the target user group and of the user needs and challenges to be addressed by the project, based on the guided interviews with the Competent National Authorities (CNAs); this includes a risk assessment of the user requirements in the context of the project.
- A review of the current state-of-the-art literature/studies on relevant EO algorithms, methods, models addressing the project needs, and the major knowledge gaps and challenges.
- A proposed technical approach to achieve the EO-integrated solution to be developed and validated; and the specifications of the test and demonstration sites that can form the basis for the Use Case studies to be conducted to demonstrate the adequacy, robustness, scalability and usefulness of the proposed solution.

The next Chapter will provide an overview of the policy review. This will be followed by a summary of the User Requirements. Chapter 4 will present a review of the state of the art technical solutions currently available to meet the different user requirements, as well as the gaps in the current methodologies. And following this the next Chapter will then describe a proposed technical approach for the current project. The report will conclude with a presentation on how the test and demonstration sites will be selected for the different commodities.

2 Policy Review

2.1 Background of EUDR

Recognizing the urgent need to address global deforestation and its significant impact on climate change, Regulation (EU) No 2023/1115 for Deforestation-Free Products Products (hereafter EUDR) was adopted by the European Parliament and the Council of the European Union on 31 May 2023 and entered into force on 29 June 2023 (Neef et al., 2024). The regulation aims to ensure that products placed on the EU market, regardless of their origin, have not contributed to deforestation or forest degradation. A key aspect of the EUDR is its focus on due diligence, and for compliance with this requirement, companies need to enhance the transparency and traceability of their commodity production. Importers and traders of designated commodities, namely *cattle*, *cocoa*, *coffee*, *oil palm, rubber, soya and wood.* See also Figure 1, which, following the planned revision on 30 June 2025, will include maize and sugarcane are required to conduct thorough assessments to verify the sustainability of their supply chains. This due diligence process involves the use of geospatial data to track the origin of products and monitor changes in forest cover.

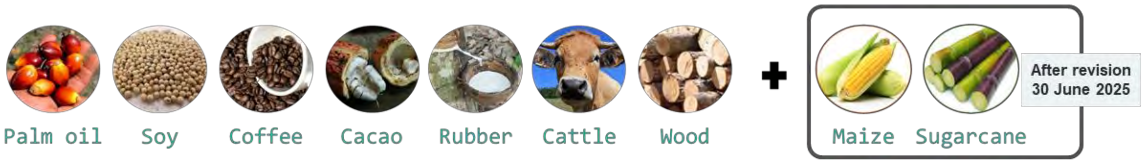

Figure 1: Products (commodities) to which the EUDR applies to (Ch. 1, art. 2, and Annex I).

The regulation (article 49) mandates that operators implement due diligence systems to ensure their products are deforestation-free and legal. These systems should include:

1. **Information Requirements**: Gathering information about the origin and supply chain of products, including geolocation coordinates of production areas.
2. **Risk Assessment**: Identifying potential risks of deforestation or illegality associated with the products.
3. **Risk Mitigation**: Taking steps to minimize or eliminate identified risks.

Provided operators confirm no significant risk of non-compliance after due diligence, they can place products on the European market.

The European Commission (EC) Benchmarking system is currently ongoing. The Competent National Authorities (CNAs) of the member states are responsible for monitoring due diligence from operators and traders. Hence, the CNAs must conduct inspections to verify operator compliance with regulations. Inspections should be prioritized based on risk levels (low, standard, high). The EUDR suggests considering the following risk criteria: rate of deforestation and degradation, rate of expansion of agricultural land, and production trends of the relevant commodities.

Due to concerns from EU member states, non-EU countries, traders, and operators about compliance difficulties, on October 2nd 2024 the European Commission has proposed a one-year delay in the implementation of the deforestation regulation. The European Parliament agreed to this postponement and other amendments in October 2024. Large operators and traders will now have to comply with the regulation by December 30, 2025, while micro and small enterprises have until June 30, 2026[1]. This extension aims to facilitate a smooth implementation of the rules without compromising the regulation's objectives.

Overall, the extension of the deadlines will not affect the WAC project timeline and schedule.

1 https://www.europarl.europa.eu/news/en/press-room/20241111IPR25340/eu-deforestation-law-parliament-wants-to-give-companies-one-more-year-to-comply

2.2 Role of EO data for monitoring and compliance

The legislation specifically encourages using EU space programs like Galileo and Copernicus to support geolocation and monitoring efforts. Figure 2 shows a simplified policy wheel, adapted to the EUDR. A policy wheel is a framework typically used to visualize the cyclical nature of policy development and implementation (Bardach and Patashnik., 2023), and here we included the potential role of EO-derived products in the different phases. Key stages of a policy wheel include the two initial stages of problem definition and proposing legislation (dark blue boxes), and implementation and evaluation (lighter blue box). Eventually, a new iteration is required with an updated problem definition. Currently, we are in the (extended) implementation phase.

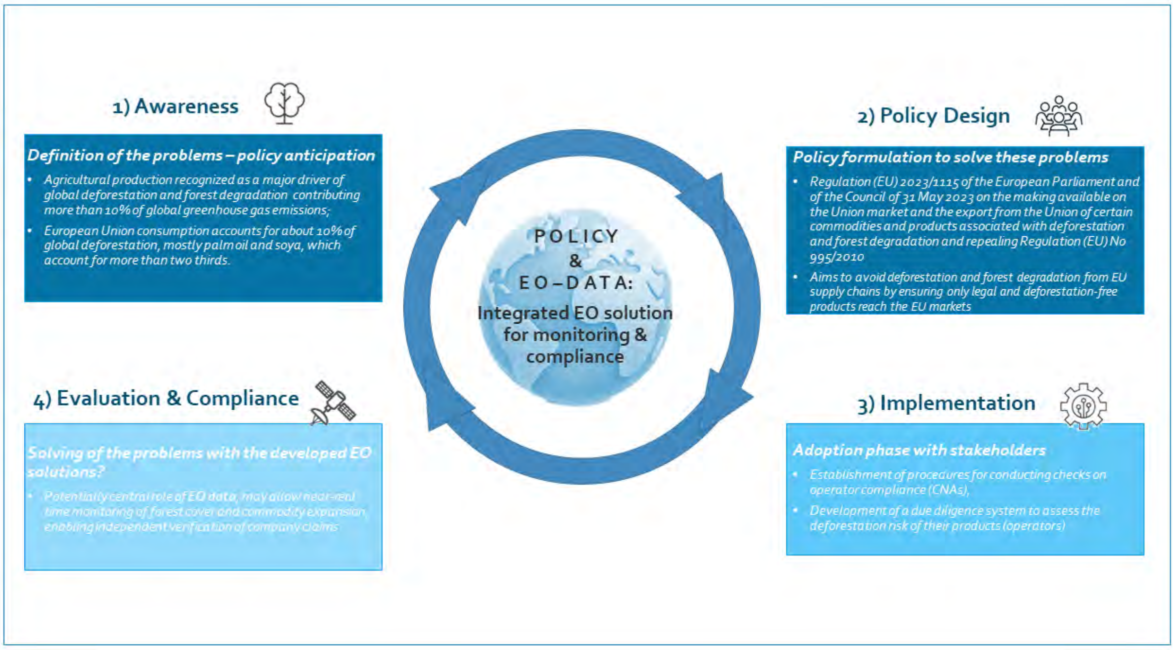

Figure 2: Simplified EO policy wheel adapted to the EUDR, incorporating EO data and product usage. The dark blue background indicates that the regulation/legislation has reached the respective stage, while the lighter colours indicate that the legislation is in the process of reaching the respective stage. Currently, the EUDR is in the implementation phase (November 2024).

Geospatial data can be utilized in various ways to support EUDR compliance. For instance, high-resolution satellite imagery can be employed to map forest cover, identify deforestation events, and monitor changes in land use over time identifying the commodities (Ribeiro et al., 2024). In addition, Geographic Information Systems (GIS) can be used to analyze and visualize geospatial data, enabling stakeholders to assess the potential impacts of production activities on forests.

Despite the increasing availability of geospatial data and advanced analytical tools, challenges remain in effectively utilizing EO-based information for EUDR compliance. Key uncertainties include the optimal resolution and frequency of satellite imagery, the appropriate methods for validating and interpreting geospatial data, and the potential costs and complexities associated with implementing robust geospatial monitoring systems (e.g., Ribeiro et al., 2024).

To obtain an overview of country preparedness, the Food and Agriculture Organization of the United Nations (FAO) undertook an assessment framework with several countries to understand how forest mapping and monitoring systems would need to evolve to best tailor to requirements (Neef et al., 2024). Discussion rounds took place in various countries in South America (e.g., Argentina, Colombia), Africa (e.g., Botswana, Cameroon), and also Indonesia (e.g., Viet Nam) providing the following major outcomes (shortened from Neef et al., 2024):

- Mapping commodities is **equally important** and potentially more difficult than mapping forests.

- For forest maps, current datasets often **do not match the required forest definition** and time frames.
- Relying on **one single dataset is problematic** for identifying forests and deforestation.
- Forest and land-use mapping can be centralized, but collecting information on plot boundaries **requires fieldwork** and it is hard to guarantee consistent high quality.
- Information on forests and deforestation must be **easily and publicly accessible**.
- Clarity is needed on how to **safeguard the privacy and property rights** of data owners, especially for smallholders.
- Farm boundary datasets could take relevance for land tenure, **creating risks of conflict among tenure holders**, which need to be carefully managed.
- Several institutions from different affiliations should **effectively collaborate to produce the needed datasets on land use, forests and deforestation.**
- Providing maps on **forests, commodities and deforestation is not a once-off task** but requires an ongoing mapping programme.

Hence, to effectively implement the EUDR, specific EO data and derived products are essential to accurately monitor forest cover, deforestation, and commodity expansion. High-resolution yearly products are useful to track long-term changes in forest cover. However, to distinguish between planted and natural forests and to assess degradation, monthly very high-resolution products are crucial. For commodity expansion monitoring, high-resolution imagery is also required to identify changes in land use, particularly in areas with significant agricultural and plantation activities: While Sentinel-1 and Sentinel-2 provide valuable data for forest monitoring, their spatial resolution may not be sufficient for EUDR compliance checks in all cases, especially for certain commodities. For example, Sentinel-2 has proven effective in detecting and mapping large-scale monoculture crops like oil palm, rubber, soy, and pasture (Descals et al., 2024; Wang et al., 2023; Song et al., 2021; Parente et al., 2024). However, for small-scale or agroforestry systems, such as coffee and cocoa, very high-resolution (VHR) imagery with a spatial resolution of less than 5 meters offers superior performance (Masolele et al., 2024). VHR data can capture intricate patterns and subtle characteristics of these small-scale systems, enabling more accurate detection and mapping. Masolele et al (2022) stated that either detailed spatial patterns (single-date VHR) and/or detailed temporal patterns (e.g., multi-date Sentinel-2) are needed for identifying land use following deforestation.

Freely accessible forest cover maps, such as the European Commission's Joint Research Centre (JRC) Global Forest Cover Map 2020 (GFC 2020) at a 10-meter spatial resolution (Bourgoin et al., 2024), provide valuable insights into the spatial extent and distribution of forests, which are crucial for tracking changes in forest cover over time. However, relying solely on single-date datasets can lead to inaccuracies, especially for specific commodities and planting systems like tree crops and agroforestry. As demonstrated by Ribeiro (2024), coffee plantations were often misclassified as forests in December 2020, the EUDR cut-off date, when analyzed using high-resolution PlanetScope imagery (see Figure 3).

Figure 3: The usage of high-resolution data to refine land cover classification: (A) The forest (in green) and non-forest map from JRC located in one of the largest coffee-producing municipalities (Patrocinio - Minas Gerais) in Brazil. (B) A PlanetScope image, from December 25, 2020, indicates that the areas inside the red squares were already coffee plantations (in courtesy of Ribero et al., 2024).

The verification and inspection of declarations made in due diligence statements by the CNAs of the EU member states, to ensure correct implementation of EUDR compliance, will require detailed, location-specific information that a global map cannot provide.

Assessing site-specific forest management practices necessitates the use of multi-temporal imagery, encompassing periods both before and after the cut-off date. This allows for a more accurate assessment of changes in forest cover and land use.

In summary, we conclude that a key requirement for EO-based data and products expressed (and expected) from diverse EUDR stakeholder groups, such as CNAs of member states, the EC, and eventually also operators and traders is the **Need for high-quality, accurate, and up-to-date EO data**, monthly and at least 10 m spatial resolution at the national level, to derive products on:

- **Commodity expansion;**
- **Occurrence of deforestation events;**
- **Baseline global forest cover 2020 but also forest cover information before and after the cut-off date.**

These products and tools will be made available to the CNAs to facilitate the monitoring of the due diligence process, particularly in tracing the origin and supply chain of designated commodities. This ensures that all operators and traders within the EU market comply with the required standards. For instance, these tools can be utilized for verifying geolocation data, conducting risk assessments, analyzing land use polygons (e.g., deforestation status, current land use for commodities), and providing guidance for verification processes.

3 User Requirements Summary

This chapter provides an in-depth analysis of the requirements of the early adopters, being the CNAs of the EU member states. As described in Chapter 0. CNAs play a crucial role in conducting due diligence on operator declarations by utilizing remote sensing data, field verification, and other tools to assess compliance with EUDR requirements. This in-depth analysis results from the feedback given by the six CNA teams (Germany, Czech Republic, France, Netherlands, Belgium and Italy) within guided interviews from September 20th to October 30th 2024. The full interview questions can be found in Annex 1. The CNA groups identified multiple specific needs in terms of EO data and tools for the implementation of an effective EUDR due diligence checking system.

The next Section provides an overall summary of all the feedback from the interviews. Further, we summarize the first Living Lab (LL#1) (3.2) and provide the risk analysis of the user requirements (3.3).

3.1 Summary Feedback from the CNAs

The guided interviews with CNA teams identified three key areas: technical requirements for EO-based solutions, selection of test and demonstration sites, and challenges related to EO-based products, including technical factors, and commodity-specific issues. These areas will be discussed in detail below.

(A) Technical requirements:

There is a consensus between all CNA groups on the need for a *2-step verification process*. The first step is to have a system, where high-risk polygons could be identified quickly and in the second step, the tool should be used for more detailed inspection-level work. Further, there was a request to monitor time stamps of land use changes, especially when related to deforestation, before and after the cut-off date on 31/12/2020 after which no deforestation should occur according to EUDR, and also assess potential changes in forest cover. They require global forest cover datasets that are aligned to EUDR and disturbance products are also welcome. Furthermore, there was the request that the tools should be user-friendly and it would be important they can be integrated with the EU systems (i.e., TRACES).

In view of forest degradation, important for CNAs is the detection of primary forest conversion into plantations. Moreover, signs of degradation, such as selective logging or clear-cutting, and distinguishing forest types between primary, and secondary forest types is of high relevance. It was also noted that the time lag between the harvest of wood and regrowth needs to be captured on time, and also needs to be linked to the Information System.

In terms of timeliness/frequency, the CNA groups want regular updates and timely access to EO data and products. It was mentioned as very important that the processing/analysis of large volumes of EO data is efficient. Regarding the specific spatial and temporal resolutions and product latency, we provide an overview in Table 1. Mainly, all CNA teams require 4-10m but for detailed inspections, a higher resolution would be preferable. For the temporal resolution, they asked for quarterly updates/monthly/yearly updates. For data processing, they want external/cloud processing and also systems which ensure data security. The system should also allow users to export analyzed data in various formats and there should be compatibility with widely used software (e.g., QGIS).

Table 1: Technical specifications given by the CNAs during the interviews, mainly in response of the question: which frequency of updates do you think is needed (yearly, monthly, NRT) and what is in your opinion the optimal spatial resolution for deforestation / agricultural mapping? VHR1: Very high resolution, VHR2: Very high resolution 2, HR: High resolution.

Countries / EO features	**Germany**	**Czech R**	**France**	**Netherlands**	**Belgium**	**Italy**[1]	**Total**
Temporal Resolution (Frequency)							
Daily							
Monthly		**X**	**X**		**X**		3
Quarterly/seasonal	**X**			**X**			2
Yearly		**X**	X				2
Spatial Resolution							
VHR1:< 1m	X						1
VHR2:>1m and ≤4 m	X			X			2
HR:>4m and ≤10 m	**X**			**X**	**X**		3
Data/Product Latency							
Real-time/On-demand	**X**	**X**					**2**
Near real-time (NRT)							

[1]missed to fill in these questions

(B) Selection of Test Sites:

We discussed in detail the selection of test sites with the CNAs and they mentioned the following specific criteria were used such as:

- **Import Volume:** countries with high import volumes of commodities linked to deforestation.
- **Accessibility:** Select areas near urban areas or roads, especially those with recent road construction which facilitates/indicates deforestation.
- **Forest Edge:** Focus on areas located at the edge of forests, where deforestation often occurs.
- **Multiple Commodity Production:** Prioritize regions with diverse commodity production, increasing the **likelihood of deforestation risks**.
- **Areas of high recent dynamics** (rather than regions where deforestation already occurred a lot)
- **Agroforestry:** While not a high priority due to lower yield potential, consider including agroforestry areas for a comprehensive assessment.

Table 2 demonstrates the preferences of all six CNAs in the form of a matrix with countries and commodities.

Table 2: Countries and commodities that CNAs mentioned as priorities: Germany (X) France (X) Netherlands (X) Czech (X) Belgium (X) Italy (X).

Countries / Commodities	Soy	Cocoa	Coffee	Oil Palm	Rubber	Wood	Beef	Total
South & North America								
Brazil	X, X		X		X	X, X, X, X	X, X	12
-Cerrado	X							
- Para	X							
Colombia			X	X				2
Argentina	X						X	2
Peru						X		1
South America general		X	X, X					3
USA	X, X					X	X	4
Africa								
Cote D'Ivoire		X, X, X			(X)	X		5
Ethiopia			X					1
Cameroon		X, X						2
Liberia		X			(X)			2
Nigeria		X						1
Ghana		X, X						3
(S-W of Komasi)		X						
Africa general				X		X		2
Asia								
Vietnam			X		X, X	X		4
Indonesia				X,X	X, X, X	X, X, X		9
(Kalimantan)						X		
Malaysia				X	X	X, X, X		5
Thailand					X			1
Russia						X (X)		2
India						X		1
S-E Asia					X			1
Asia general						X		1
Europe								
Scandinavia (Sweden & Finland)						X		1
Europe general								
All producing countries			X					1
Total	**7**	**11**	**7**	**4**	**11**	**21**	**4**	**66**

(C) Challenges:

Multiple challenges were noted by the CNA teams for the upcoming implementation period. Regarding **EO-based monitoring**, for instance, distinguishing forest types, agroforestry and small-scale farming systems was mentioned several times as a difficulty. Also, identifying and monitoring small-holder producers, particularly in remote areas with EO data is challenging and VHR would be needed. Moreover, mapping challenges in tropical moist areas where commodities mostly grow are expected.

Other constraints are of **technical nature**. Some CNA groups do not possess experience in creating IT systems with satellite data and geo-information aspects. One of the main challenges mentioned will be that within the due diligence, operators will provide a massive amount of data and a tool is needed that allows checking annually if the yield of that plot is realistic. It is feared that geolocation data provided by the operators may be used more than once. **Table 3** lists the **commodity-specific challenges.**

Table 3: Commodity-specific challenges mentioned by the CNAs:

 Rubber	*Rubber*: it was noted that with rubber, tracking the origin could be challenging due to mixed sources (plantations and smallholders).
 Palm oil	*Palm oil:* due to rapid expansion in the crop, there is often illegal deforestation. For both *oil palm* and *rubber,* the new EU Guidance document indicates certain species only will be affected.
  Coffee Cacao	*Coffee* and *cocoa:* share the challenge of mountainous terrain agroforestry systems.
 Wood	*Wood:* there will be challenges in the correct assessment of degradation.
 Soy	*Soy:* in the due diligence, no date is given in the declaration, and since *soy* is an annual crop, it might be difficult to assess. Also, there is the challenge of soy coming from different sources in a mixed shipment.
 Cattle	*Cattle:* the main requirement in EUDR is the need for point data which is related to the farm; however, as cattle move around and go to slaughterhouses at different locations to their farms, different areas need to be checked.

Besides the commodities itself, highly processed (derived) products and products with a long supply chain will be challenging.

Some challenges related to data quality/verification were noted as the need to ensure the accuracy of the EO data, the need to develop robust methods and the establishment of clear legal frameworks for EO-based monitoring.

Finally, it should be emphasized that the development of the EO-based monitoring system will follow an iterative approach, closely adhering to agile principles and involving CNAs in different development stages. This collaborative process will ensure continuous refinement and alignment with user requirements and the underlying policy framework, ultimately delivering an optimal EO-based solution.

3.2 Outcome summary of the LL#1

3.2.1 Overview

The LL#1 took place as a hybrid meeting on Nov. 7th 2024, with the option either to attend physically in Brussels or join online. Overall, we had a very satisfying attendance of 25 participants, 13 physical and 12 online (see names in Annex 2 of the minutes document), covering the whole consortium, the six CNA groups as well as some participants from the Stakeholder Engagement Facility (SEF). The meeting was chaired by Christophe Sannier (GAF) and lasted from 9:00 to 14:30. See Annex 2 for Agenda and Workshop Report.

The main objectives of LL#1 were to review in-depth the idea of the project; provide an overview of the CNA requirements that have been collected in the two months since the project KO, discuss the geographical areas for test/demo sites; present a concept for the technical approach, present additional project activities where Consortium need CNA support and finally organize the collection of in situ data. In the following we will detail the feedback for user engagement summary, and give an overview on presented plans for the proposed integrated EO solutions. Some additional feedback was provided by the CNAs during the Workshop which is noted herewith. The full Minutes of the Workshop however can be found in Annex 2.

3.2.2 Feedback regarding user requirements & test sites

There was some relevant feedback given from the CNAs following the presentation of the user engagements. The discussion centred on the selection of test sites and the proposed two-step approach for data assessment. Participants suggested including an EU country for the forest sector, specifically considering agroforestry and potentially Ghana's Sui Forest Reserve.

The two-step approach involves an initial automated assessment to identify high-risk cases, followed by a more in-depth inspection using VHR data. The goal is to verify declarations, ensure correct commodity identification, and potentially assess biomass extraction for plausibility checks. Participants emphasized the importance of considering multiple geolocations per declaration and using buffer zones to account for adjacent land.

The discussion highlighted the challenges of handling a large volume of declarations, particularly for countries like the Netherlands. Interoperability between systems like TRACES and getting a red flag for a declaration was seen as crucial for efficient processing.

The issue of false positives of the GFC 2020 in agroforestry was raised, especially when dealing with indigenous communities. The need for accurate and non-disruptive monitoring methods was emphasized. CNAs offered to share insights from their upcoming field visits to Brazil, which would help address these challenges and develop more effective monitoring strategies.

The German CNA team discussed the ongoing work on information systems and control plans, emphasizing the importance of incorporating risk criteria like proximity to forests. They also shared concerns about the accuracy of the GFC 2020 map, particularly for cocoa and coffee in South America. The groups agreed that the JRC might need to provide more guidance on accuracy issues of the GFC 2020, especially regarding false positives in cocoa and coffee areas.

The CNA of the six member states fully agreed upon their next involvement steps, as detailed also in **D.1.1.**

3.3 Risk analysis of the user requirements

As the CNAs presented various requirements that come with challenges it’s important at this stage to conduct a risk assessment for these issues. The proposal also had conducted a risk assessment for the project and some of these risks are still relevant and addressed herewith. See Table 4. The various problem types were assessed in terms of the risk levels according to the impact on the project if it occurs, the probability of occurrence and the proposed mitigation measures.

Table 4: Risk analysis of the user requirements

Category	Potential Problem Description	Description of Impact on Project	Impact	Probability	Mitigation Strategy
Organisational Issues	CNA engagement and feedback irregular and/or weak.	Several activities in the Tasks requires strong stakeholder engagement and therefore this would have a negative impact on overall fulfilment of these Tasks.	Drastic	Unlikely	The CNA collaboration was supported at project inception with support from ESA. The guided interviews were a first successful example of the readiness of the CNAs for the collaboration. Additionally, the Project Co-ordinator presented the various Tasks which require their support in the first Living Lab Workshop to raise their awareness. Finally, the Consortium have experience in stakeholder engagement and will ensure that CNAs are kept up to date on project progress, and that any issues of concern can be quickly resolved. A main mitigation measure is that there is an urgent need for technical solutions by the CNAs so they have a vested interest to be involved in the development of the project activities.
	Products and services do not meet stakeholder expectations.	Stakeholder validation would be negative and utility of products for work practices would be nominal.	Large	Unlikely	The envisaged Task 1 and 3-5 have the objective to liaise closely with the CNAs and other stakeholders on the product/EO solution requirements and to come to a consensus on product specifications and product utility. If these activities are implemented in a proficient manner via the Living Labs, the stakeholder Workshops, and promotion activities it is very unlikely that the stakeholders will be dis-satisfied with the final products.
Technical Issues	Degradation assessment.	The EUDR requires forest degradation assessment especially when related to wood products.	Large	Likely	Approaches such as canopy disturbance and Intact and non Intact forest assessment can be used as proxies for the degradation mapping. The Consortium, have experience and tools for these approaches which can be implemented.
	Agroforestry systems detection.	Majority of cocoa/coffee farms in Africa and some in South America are grown by small holder farms using agroforestry systems.	Drastic	Likely	The need for in-situ data which the Consortium has for these 2 commodities in different countries will support the improved detection and classification of the agroforestry practices.
	Classification of forests for Degradation assessment.	To further support forest degradation, forest classification should be further refined into separate categories of primary and secondary forest.	Large	Unlikely	The use of the Tree Canopy Disturbance Model (TCDM) provided by GAF can be used as a proxy indicator for degradation assessment as well as other methods such as Intact/non-Intact Forest assessment.
	Location of cattle during their life cycle.	Cattle: the main requirement in EUDR is the need for point data which is related to the farm; however, as cattle move around and go to slaughterhouses at different locations to their farms, different areas need to be checked.	Large	Likely	The Consortium will have access to the Brazilian Cadastro Ambiental Rural which has systematically collected all farm property data for the country; the database has both the geolocation information as well as commodities grown including that of cattle. The CAR is a public domain dataset. Other public datasets such as that from the TRASE dataset will also be used.
	Availability of cloud free EO data.	Due to the tropical climate in some of the countries where the test and demo sites will be located, it has already been noted that many areas are cloud covered for most of the year. This impacts the ability to use optical EO data for provision of some of the products.	Large	Likely	The problem with cloud cover can be addressed using radar data whenever possible; furthermore, the use of dense time series optical data (from e.g. Sentinel-2, Landsat) which with its frequent acquisition dates will provide more opportunities for accessing cloud free data and compensate cloud cover related data gaps in the data processing. Further, alternate satellite data will be evaluated for their suitability.
	Limited access to in-situ country data.	Would be a problem for test and demo site implementation for small scale plots and for overall validation of results.	Large	Unlikely	The Consortium partners Terrasphere, Meridia and WU have combined their different in-country datasets and there is both an adequate volume and wide geographic spread of the in country data.

4 Review of State-of-the-Art Technical Approaches

4.1 Overview

The monitoring and mapping of commodity crops such as rubber, oil palm, cocoa, coffee, pasture, timber, and soy have garnered significant attention in recent years due to their critical role in global trade, sustainability, and compliance with frameworks such as the European Union Deforestation Regulation (EUDR), European Commission (2024). Advances in Earth Observation (EO) technologies and machine learning (ML) models have enabled the generation of high-resolution maps of these crops, albeit with varying thematic, spatial, and temporal details. Many efforts, however are based on small scales or regional level (Wang et al., 2023, Kalischek et al., 2023). Maps of commodity crops at global scale have been attributed to lower accuracies. The accuracy for classifying commodity crops is largely affected by the heterogeneous nature of climate, geographical conditions where these crops are growing. The difficulty stems partially from similarities and differences, both in spectral and spatial aspects, between commodity crop farming practices. Commodity crops over large scales are spectrally heterogeneous, with substantial variation in spectral signatures across space, time, elevation, soil types, forest types, and disturbance intensities. These similarities and differences, along with variation in spectral reflectance and phenology across geographic regions, as well as persistent tropical cloudiness, make it difficult to distinguish commodity crops using satellite imagery consistently. This has an implication on the choice of methods, and its performance on identifying and mapping respective commodity crops. Therefore, the same approach is unlikely to be applicable to all commodities and specificities of each commodity in terms of landscape and spectro-temporal characteristics as well as different agricultural practices will need to be accounted for.

4.2 State-of-the-Art EO Methods for Commodity Crop Mapping

4.2.1 Oil Palm

Currently there exist several commodity crops products based on several studies, however at varying thematic, spatial, temporal detail and accuracies. Dong et al. (2020) used Convolution Neural Networks (CNNs) to classify Level-18 Google Earth high-resolution images, mapping oil palm plantations across Malaysia. The study highlighted the utility of localized spatial features extracted by CNNs in distinguishing oil palm from other land covers. Descals et al., (2024) mapped oil palm plantations at 10 m resolution using Sentinel-1 images and CNN across the globe. In their mapping effort they achieved a producer's and user's accuracy of 91.0 ± 2.5 % and 91.8 ± 1.2 % for industrial plantations and 71 % and 72 % for smallholders. Although this study achieved high accuracy on already established industrial oil palm plantation, challenges still exists, especially in identifying young plantations, smallholders oil palm and open canopy oil palm where low accuracies are observed.

4.2.2 Rubber

A study by Somching et al. (2020) used Landsat time-series data and ML algorithms to identify the establishment year of rubber plantations in Thailand, providing insights into temporal changes and plantation dynamics. Wang et al., (2023), produced a map of rubber plantations for Southeast Asia at 10 m resolution using Sentinel-2 images and random forest machine learning algorithm (ML). They achieved, user's accuracy of 99%, and producer's accuracy 95%, however the accuracy dropped to 0.57 when adjusted by the area proportion of the class. The main challenge they faced in this study is identifying young rubber plantations and confusion of rubber with oil palm plantations. Likewise there were also challenges in upscaling the mapping of rubber to regional scale, caused by lack of distinctive seasonal patterns near the equator resulting in greater variability of accuracies in identifying rubber across regions.

4.2.3 Cocoa

Another study by Abu et al., (2021) utilized a multi-feature Random Forest (RF) algorithm to map cocoa plantations in Ivory coast and Ghana. The Random forest model was trained on multi-temporal stack of Sentinel-1 and Sentinel-2 images data to classify cocoa trees, enabling accurate mapping of cocoa cultivation areas with 82% producer's and 62% user's accuracy . This is also complemented with a study by Kalischek et al., (2023) who mapped cocoa plantations in Ivory coast and Ghana using CNN and Sentinel-2 images with an accuracy of 88%. While the model from this study shows promising results, adapting this model to other regions with different cultivation practices may be challenging. Likewise the dependency on optical data limits its application in regions with high cloud persistence.

4.2.4 Soybean

Maps for soybean, are also available from other studies, for example soy data by Song et al., (2021) employs satellite data and decision tree ensemble model to map soybean fields across South America. The decision tree ensemble model is trained on multi-temporal Landsat and Modis data to identify soybean crops, facilitating the monitoring of soybean expansion and land use changes in the region, achieving the overall accuracies of 96%, 94% and 96%, for 2017, 2018 and 2019 respectively. Li et al. (2023) used monthly composites of all Sentinel-2 surface reflectance bands in a Random Forest approach to produce the first wall-to-wall map of maize and soybean at 10 m resolution for China, for the year 2019.

4.2.5 Coffee

A research by Maskell et al., (2021) utilized Sentinel-1 and Sentinel-2 satellite imagery to map coffee plantations in Vietnam. They used texture and temporal statistics information to delineate coffee fields accurately with an overall accuracy of 89% user accuracies of 65, 56, 71% when the accuracies are computed based on three coffee production systems namely sun coffee, intercropped coffee and newly planted coffee, respectively. It is important to note that the maps are produced at small scale as currently there is no regional or global scale maps of coffee farms.

4.2.6 Pasture

An effort by Parente et al., (2024), employs Landsat-8 data and machine learning techniques to map pasture land expansion worldwide. The Random forest model is trained on Landsat imagery to detect changes in land cover associated with pasture establishment with a balanced user's and producer's accuracy, resulting in F1 score of 64% and 75% for cultivated and natural/semi-natural grassland, respectively. The study emphasized the simplicity of RF but noted its limitations in incorporating spatial and temporal information.

4.2.7 Timber (Logging)

Dalagnol et al. (2023) & Hethcoat et al. (2021) utilized RF models to map forest degradation due to logging in the Brazilian Amazon. These studies used Planet-NICFI and Sentinel-1 imagery, respectively, focusing on logging-associated forest changes.

4.3 Comprehensive Multi-Class Mapping

Now, most of the maps from the above studies either focus on a single commodity crop class or are produced at a temporal resolution not sufficiently detailed to make them useful for timely applications. Likewise deriving deforestation statistics from a single commodity prediction, can underestimate the influence of other commodity crops on deforestation. For example, currently most of the soy expansion in Brazil occurs in what was previously pastureland, thus indirectly influencing deforestation by pasture, (Song et al.,2021). Holistic approaches that track land use across all commodities coupled with vegetation monitoring are required to maintain critical and accurate reporting on deforestation drivers. A study by Masolele et al., (2024), maps commodity crops associated with deforestation in Africa with an accuracy of 84% from the year 2001 to 2020, and using CNN algorithms and 5 m planet-NICFI images, making it easier to derive commodity crops contributing to deforestation at local scale. The study also includes more thematic details covering the 7 classes currently included in the EUDR framework. A similar study by Masolele et al., (2022) produced maps of land use following deforestation in Ethiopia (and included commodity crops highlighted in the EUDR such as coffee, a main commodity crop in Ethiopia).

Most if not all multi-class crop maps to date have been produced using a local to regionally finetuned approach, where local in-situ reference data is used to train a dedicated machine learning model based on manually selected classification features originating from time series of one or multiple satellite data sources. Such models are typically poorly generalizable across space and time, making them unsuitable for an operational mapping system which inherently needs to be able to cover multiple years and regions. Most popular satellite data include Landsat, Sentinel-2 and Sentinel-1 data. A pixel-based Random Forest approach remains the most popular approach due to its low requirement for in-situ training labels and relatively good performance. Building on top of this approach, but focusing particularly on temporal and spatial generalization, Van Tricht et al. (2023) created the first global classification models for cereals and maize in the WorldCereal project. They used a pixel-based CatBoost classifier on top of manually selected classification features from Sentinel-1 and Sentinel-2 mainly. In case more training data is available, researchers typically turn to more advanced classification approaches, like CNN (taking into account spatial context), or more recently transformer-based models, which leverage the powerful capabilities of deep learning to let the algorithm discover the most important and relevant features in the satellite data for the task at hand, eliminating the need for manual feature selection (Broekx, 2023). VITO recently used the latter approach to produce a comprehensive multi-class crop type map for the entire European continent and multiple years in light of the new Copernicus High Resolution Vegetated Land Cover Characteristics service.

4.4 Existing Toolbox and Workflows

Although crop type mapping based on EO data has been around for many years, standardization of the required tools and accessibility of those tools to non-expert users remains low. Up till recently, creating crop type maps, especially at large spatial scales, required manually setting up a dedicated processing cluster, preferably in a cloud-based processing environment (e.g. using Amazon Web Services, or AWS, and Google Earth Engine), and then constructing and deploying dedicated EO data processing pipelines, including EO data download, pre-processing, feature computation, model training, model inference and post-processing.

In the framework of ESA Sen2Agri project, the Sen2Agri toolbox (Université Catolique de Louvain, 2020) was created, representing the first standalone system allowing end-to-end processing of satellite data into crop type maps and other relevant products. The Sen2Agri system has been deployed on AWS and Food Security TEP cloud environments, with the intention to help countries create national crop type maps, yet uptake of the system by the agricultural community remains limited.

Most recently, the ESA WorldCereal project has launched the first version of its generic, cloud-based crop mapping system (VITO Remote Sensing, 2024). This system represents an open and easily accessible system and encompasses all steps related to the production of crop type maps, starting from in-situ data harmonization, management and sharing in its dedicated Reference Data Module (https://ewoc-rdm-ui.iiasa.ac.at/).

The processing module responsible for map generation will be implemented within the Copernicus Data Space Ecosystem (CDSE) using the OpenEO API. OpenEO offers a range of features that make it highly suitable for generating crop type maps at scale.

At its core, OpenEO is a data cube-centric approach for working with EO data. It has been under continuous development since 2018, starting as an initiative funded by the H2020 program. Over time, OpenEO has evolved into a user-friendly ecosystem featuring graphical interfaces and programming libraries, all of which connect seamlessly to the OpenEO API. Operational deployment has been achieved through the ESA-funded OpenEO Platform project, which is currently integrated into CDSE, with funding secured until at least 2028.

The platform is supported by multiple cloud providers and serves as a key component of the CDSE infrastructure. OpenEO provides standardized access to diverse EO dataset collections and offers extendable functionalities designed for compatibility and integration with the broader EO data landscape. Its federated architecture connects to various interfaces, including EODC, Terrascope, CreoDIAS, OTC, and CDSE. This versatile framework has been successfully applied to numerous projects and use cases, such as land cover mapping. OpenEO workflows can run on any system equipped with an OpenEO backend, ensuring scalability and adaptability.

A standout feature of OpenEO is its ability to blend predefined and user-defined functions, enabling backend optimization while granting users access to extensive libraries and tools. The platform connects to over 100 EO datasets, or "collections," through various backends. If a required dataset is not yet implemented, it can be added as a custom collection. OpenEO also ensures full access to data available on the CDSE platform.

The platform utilizes the open-source Geotrellis implementation, powered by Apache Spark for distributed processing. This enables it to perform complex operations, such as merging data cubes and applying functions across complete time series for individual pixels.

OpenEO adheres to the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles for managing the data lifecycle. It employs the STAC metadata standard to describe input and output datasets. Built-in support for FAIR-compliant metadata simplifies project workflows and enhances the traceability of all generated data in a standardized format. The platform also supports multiple output formats, including GeoTIFF, netCDF, and PNG, enabling seamless integration with catalogues and data viewers.

If a workflow cannot be fully converted to OpenEO, individual processing steps can be packaged as microservices within cloud-based Docker containers. This approach has already been applied successfully in other services for generating tree cover density (TCD).

4.5 Open processing environments

The main data processing system proposed for this project aims to efficiently handle the data processing using state-of-the-art free and open-source software. Our first choice for processing will be the CDSE with its existing and thoroughly tested processing capabilities (powered by VITO and Sinergise among others). Our intended solution will be fully based on open source solutions, cloud computing and based 100% on using Copernicus access and exploitation infrastructures. The complete archives of all major satellite data sources to be used in this project (Sentinel-1 and Sentinel-2 data) are readily accessible on CDSE, representing a huge benefit over other potential cloud-based processing environments, as it is generally cheaper to process the data where it resides.

As a redundancy measure the infrastructure can be deployed to an alternative provider if needed. In fact, VITO has successfully demonstrated to deploy its operational services on the WEkEO DIAS (HR VPP, HRL VLCC), OTC (LCFM), as well as other hyperscalers (WorldCover & WorldCereal).

4.6 Knowledge Gaps and Challenges

Based on the review of user requirements and current state of the art approached the following gaps and challenges have been identified:

- Thematic Gaps
    - **Single-Crop Focus:** Many studies focus on individual commodity crops, limiting their utility for broader applications that require multi-class mapping aligned with frameworks like the EUDR.

    - **Limited Attribution to Deforestation:** Few studies explicitly link commodity crop mapping to deforestation, leaving a gap in understanding the drivers of land-use change.
- Spatial Gaps
    - **Resolution Constraints:** Most EO studies utilize data at 10 m or lower spatial resolution (e.g., Sentinel-1, Sentinel-2), which may not capture local-scale variability required for farm-level monitoring.
    - **Underrepresentation of Agroforestry Systems:** Crops like coffee and cocoa, often grown in agroforestry systems, are challenging to map due to mixed signals from overstory canopy and understory crops.
- Temporal Gaps
    - **Inadequate Temporal Monitoring:** Few studies leverage time-series data to monitor crop dynamics, establishment years, or seasonal variations effectively.
    - **Change Detection**: Limited research exists on mapping the temporal expansion or reduction of commodity crop plantations.
- Methodological Challenges
    - **Feature Engineering:** Methods like RF require extensive feature engineering to integrate spatial and temporal information effectively.
    - **Model Generalization:** CNNs, while powerful, require large, diverse datasets to generalize well across geographies, leading to challenges in scaling mapping efforts globally. Most crop type and commodity mapping projects currently train localized models based on local to regional datasets, resulting in poor transferability across space and time.
    - **Data Accessibility:** High-resolution imagery like Planet-NICFI provides detailed mapping but often involves licensing and cost constraints, limiting its broader adoption.
    - **Model integration:** Integrating state-of-the-art machine learning (ML) models into openEO workflows is an area with limited prior experience. To ensure successful implementation, dedicated efforts will be required to validate the outputs generated by these models. This step is critical to maintain the accuracy and reliability of the workflows.
    - **Data Privacy:** openEO, as an open software API, is designed to align with FAIR data principles. While it can integrate private data into its workflows, special attention must be given to addressing the privacy requirements of the CNAs. Dedicated efforts will be necessary to ensure compliance with privacy standards and safeguard sensitive data throughout the workflow.

Current EO methodologies for mapping commodity crops have advanced significantly, leveraging ML and deep learning approaches. However, major gaps in thematic coverage, spatial resolution, temporal monitoring, and scalability persist. This project addresses these gaps by leveraging various open source satellite imagery (e.g., Sentinel-1/2, and ancillary data) to develop advanced ML algorithms that are able to produce detailed maps of commodity crops, contributing to improved monitoring and sustainable land-use practices under the EUDR framework. To achieve this task and based on literature review, we will conduct a comprehensive cross-comparison of state-of-the-art satellite data (Sentinel-1, Sentinel-2) and advanced machine learning approaches (e.g., Random Forest, CNN, Temporal CNN, LSTM, and Vision Transformers) for mapping the seven EUDR commodity crops. Benchmarking experiments and cross-assessments will be conducted across diverse test sites to critically evaluate the strengths and limitations of candidate models and data modalities. The project will focus on testing single- and multi-date imagery, integrating multi-resolution data, and addressing both multiclass and binary classification tasks. Additionally, we will assess the performance of models across varying regional scales, farming systems, computational efficiency (training, testing, and prediction costs), and field sizes. These efforts aim to identify optimal parameters for accurate, scalable, and operational commodity crop mapping solutions.

Qualitative and quantitative assessment criteria will be defined to evaluate the models' performance, operational efficiency, and user relevance, ensuring alignment with stakeholder requirements and the EUDR framework. Ultimately, this project aspires to deliver EO-integrated solutions capable of producing high-accuracy, scalable maps of commodity crops, contributing to sustainable land-use practices and effective monitoring of deforestation-related cash crops. By leveraging open-source satellite imagery and state-of-the-art algorithms, we aim to set a benchmark for EO-based commodity crop mapping globally, ensuring both scientific rigor and practical applicability in supporting global sustainability goals.

The integration of OpenEO into an operational service marks a significant milestone in leveraging open-source technology for environmental monitoring. For national competent authorities tasked with assessing compliance with regulations like the EUDR, this approach offers a powerful and scalable solution. OpenEO's open-source nature ensures transparency, adaptability, and alignment with FAIR data principles, which are crucial for building trust and facilitating collaboration across member states.

This implementation represents the first instance of coupling OpenEO with such a high-level operational service, setting a precedent for how open-source tools can enhance the effectiveness of EU-wide monitoring frameworks. Furthermore, by taking innovative steps to integrate advanced machine learning workflows into OpenEO through the use of User-Defined Functions (UDFs), we are unlocking new possibilities for automating complex analyses, improving accuracy, and scaling these solutions to address diverse and evolving challenges.

5 Proposed Technical Approach for Project

The present chapter is a preliminary attempt to translate the requirements as expressed from the CNAs into some technical requirements for the WAC prototype system as required in the SoW.

As highlighted from the CNA feedback in Section 3.1, a 2-step approach is necessary to best meet the requirements of the CNAs: a rapid potential non-compliance detection system followed by an evidence-based verification system.

5.1 Potential non-compliance detection system

The purpose of this first step is to sieve through a large number of Due Diligence Submissions (DDS) to identify those with a large number of potential non-compliances. A set of criteria and a threshold-based approach can be applied to identify DDS that warrant further more detailed investigation. The purpose of the first step is to provide a system that is as much as automated as possible to make it possible to process large amounts of DDS in a short time.

The main functional characteristics of the approach should be as follows:

- Perform a data integrity check including topology and basic verification against available land cover data to detect inconsistencies (e.g. water presence).
- Apply a buffer (to be determined based on tests) around the submitted polygons and point dataset to produce a relevant area of interest (AoI) for each submitted DDS.
- Identify if any forest was present before 31 December 2020 within the AoI.
- Apply a commodity classification model to the identified parcels.
- Perform a plausibility check of the declared volume against the commodity area detected for each DDS to assess if this falls within accepted yield figures.
- Analyse discrepancies between the DDS and analysis results as separate attributes in the submitted polygons / point dataset.
- Provide an overall ranking based on feedback from CNAs.

The technical specifications would be as follows:

- Spatial resolution: The main limitation is the use of open data therefore this would be 10m from Sentinel-2 or, if available / applicable (i.e. a suitable licence allows for it), 4m from NICFI initiative. No Minimum Mapping Unit or Minimum Mapping Width would apply other than that applicable to the Forest Definition.
- Temporal resolution: analysis should be contemporary to the DDS.
- Thematic accuracy: a high level of thematic accuracy should be applicable for each of the processed DDS with a target user/producer accuracy of 85-90% for both producer and user accuracy.
- The tool should be made available in a modular form as standalone cloud-based applications to facilitate the future integration in CNAs own systems (if required) and integrated as a web-based platform for demonstration purposes during the project.

5.2 Evidence-based verification system

The main purpose of the system would be to take the output of the potential non-compliance detection system and perform additional verification with the aim to confirm the non-compliance and provide evidence that could potentially be presented to court.

Technically the system will mostly consist of a visual interface to allow comparison of DDS with available datasets and the ability to order VHR imagery to confirm or not non-compliance issues. The system will also allow the analysis of any supporting data provided as part of the DDS (e.g; imagery, field photos, and map layers of indigenous lands / protected areas if available).

6 Test and Demonstration Site Selection

6.1 Overview of the approach

As an analytical framework and to aggregate the data, aid in site-selection and obfuscate sensitive in-situ data the H3 gridding system was used. The H3 gridding system is developed within Uber (see Figure 4 and also https://h3geo.org/) as an open-source, discrete, global and hexagonal grid system available for multiple resolution levels. The main advantage of the H3 grid system is the provision of a common global analytical framework minimizing geometric distortion and allowing the integration of numerous datasets.

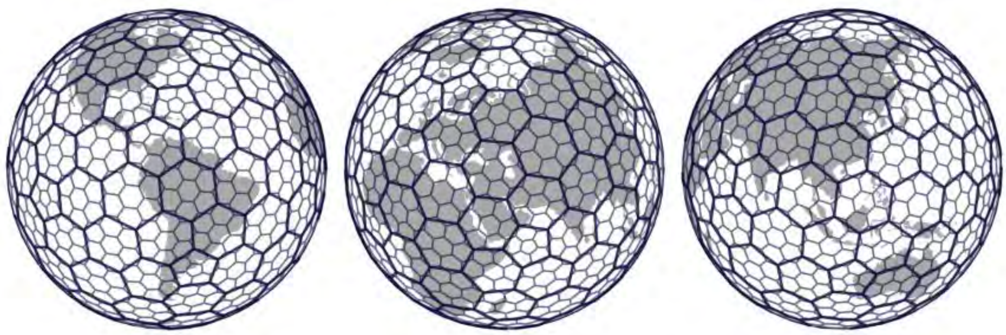

Figure 4: Global coverage with the H3 hexagonal grid system (source: https://www.uber.com/en-DE/blog/h3/)

Each H3 level 5 grid cell covers an area of about 125km² corresponding to the area of the test sites to be used as part of the WAC project. The availability of in situ data was reported in counts (points data) and hectarage (polygons) for each H3 level 5 grid and for each of the agrocommodities. No additional information on the spatial distribution of the in-situ data within the sites was provided. H3 grid cells were filtered per commodities and sites with no data were removed. The resulting dataset was then subsequently used to aggregate other data sources required as part of the test sites selection process.

The sites were defined using H3 grid level 5 cells. The hexagons contain information related to the abundance of in situ data for each of the EUDR commodities as well as information from other data sources. The selection of test sites was based on integrating four predefined criteria: (i) risk of deforestation, (ii) landscape characteristics, (iii) presence of the agricultural commodities (agrocommodities) and (iv) availability of in situ data. Firstly, each of the criteria was assessed. To obtain the risk of deforestation, the percentage of tree cover, the deforestation rate and the density of the road network were combined, and the sites were grouped into three levels of deforestation risk (Low – Medium – High). To characterize the landscape, the level of fragmentation of the land cover was estimated and grouped in each site, and the sites were grouped in three classes (Low – Medium – High) based on their level of complexity; the high value corresponding to high level of fragmentation, etc. To obtain an appreciation of the presence of the agrocommodities, several sources including land cover maps, reports and statistics on production and/or exportation were used to quantify the abundance of each commodity per site. Secondly, the abundance of in situ data was based on the counts (points data) and the hectarage (area), where the sites with higher counts and/or higher hectarage of in situ data for a given commodity was labelled as high. Secondly, the distribution of test sites per country, resulting from the discussion with CNAs, was used as reference to the number and location of sites for each commodity. Finally, candidate test sites for each of the agrocommodities were selected by combining the selected criteria in different category levels.

The data availability reported in the attribute table of the sites either as counts (points data) and hectarage (polygons) for each of the agrocommodities was used as the filter for the final selection of test sites. Sites were ranked by their volume of in situ data within a given category (level of deforestation risk and landscape complexity for a given commodity), and selected from where the volume of in situ data was the highest to lowest, based on the number of sites needed as specified in Table 5.

Finally the demonstration sites will be based on the outcome of the test site implementation ensuring that priorities from CNAs are carefully considered and a minimum of 4 countries are included per commodities as stated in the SoW

6.2 Distribution of sites per country

The test sites are equally distributed between the seven agrocommodities, with 10 test sites per agrocommodity. 33 test sites are located in South and North America, 20 in Africa and 17 in Asia. The number of test sites per agrocommodities and per country is summarized in Table 4. Countries with largest number of test sites include Brazil (12 sites), Cote d'Ivoire (8 sites) and Indonesia (6 sites).

Table 5: Distribution of the test site in countries of interest

Pays	Soy	Cocoa	Coffee	Oil palm	Rubber	Wood	Beef	Nr of sites
South & North America								
Brazil	4		2			2	4	12
Columbia			3	2				5
Argentina	3						3	6
Peru						1		1
USA	3					1	3	7
Ecuador		2						2
Africa								
Cote d'Ivoire		3		2	2	1		8
Ethiopia			3					3
Cameroon		1		1				2
Liberia					2			2
Nigeria		1						1
Ghana		2						2
Gabon						2		2
Asia								
Vietnam			2		2	1		5
Indonesia		1		2	2	1		6
Malaysia				2	1	1		4
Thailand				1	1			2
total	10	10	10	10	10	10	10	70

For a site to be selected as test site, certain requirements were specified such as the availability of in situ data, area associated with high land cover dynamic, production level, accessibility, forest edge, multiple commodity production and agroforestry. These requirements were grouped into to generate the deforestation risk which integrates accessibility, forest edge and land cover dynamic; the landscape complexity and the location of agrocommodities which gave an estimation of the production level per site.

6.3 Data description and processing

6.3.1 Deforestation Risk

6.3.1.1 Percentage of forest cover

The percentage of forest cover was used as a proxy for forest edge. The Join Research Centre (JRC) Global Forest Cover (GFC) 2020 from Bourgoin et al. (2024) was selected for this step. It provides a global analysis of forested areas and has the percentage of forest cover from 0 to 100% at 30 m spatial resolution. The mean forest cover percentage for each site was generated with zonal statistics. The value range between the test sites was scaled between 0 and 100 using the formula below.

$$Scaled\ value = \frac{(\text{actual value} - \text{min(value)}}{(\max(value) - \min(value))} \text{x} 100$$

6.3.1.2 Deforestation rate

Information about the deforestation rate was extracted from the high-Resolution global maps of 21st century forest cover change published by Hansen et al. (2013). The Hansen Global Forest change results from time series analysis of Landsat images in characterizing global forest extent and change between 2000 and 2023. The forest loss between 2020 and 2022 was used to determine the global rate of deforestation. The values corresponding to the deforestation rate over the period of observation ranged between 0 to 14.59% at 30 m spatial resolution. The zonal statistics analysis was carried to determine the average deforestation rate in each site, and the values were scaled between 0 to 100.

6.3.1.3 Accessibility

The total length of road was used as a proxy to quantify the accessibility in each site. Open Street Map (OSM) was used as data source to extract the total extent of road network in each site. The values corresponding to the total length of roads within each grid cell were scaled from 0 to 100 for each country (considering that the development of the road network different across countries).

6.3.1.4 Combined criterion

To determine the deforestation risk, a deforestation score was calculated for each site. For that, the scaled parameters previously calculated were averaged. The deforestation score across the sites was grouped into three level of deforestation risk (Low – Medium – High) based on the natural breaks in its distribution. Figure 4 summarizes to get the deforestation risk per site. From left to right, there is the scaled values of the forest cover, the scaled values of the deforestation rate, the scaled values for accessibility in sites located in Argentina

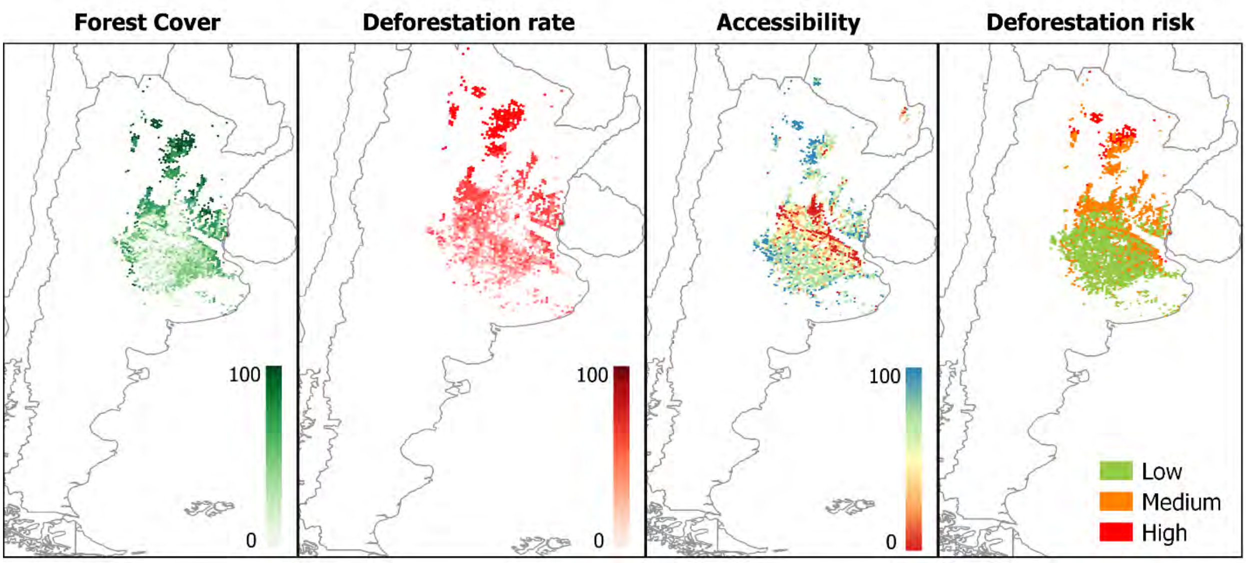

Figure 5: Illustration of the integration of forest cover, deforestation rate and accessibility to generate the deforestation risk in each site (example for Argentina).

6.3.2 Landscape complexity

The landscape fragmentation was the proxy used to describe landscape complexity in the sites. It quantifies the amount of isolated land cover classes within an area and the higher the value, the more fragmented/complex is the landscape in that site. The land cover information was extracted from the spatial analysis of the ESA World Cover map (2020). It is a global land cover map produced by the European Space Agency based on Sentinel-1 and Sentinel-2 imagery data. It provides a detailed land cover classification with 11 distinct land cover classes at 10 m resolution. The spatial analysis consisted of grouping together (clumping) neighbouring patches from the same class, and the total number of clusters was estimated in each site. The values across sites were scaled between 0 to 100, and grouped into three levels (Low – Medium – High) following the natural breaks of the distribution (Jenks, 1967) designed to determine the best arrangement of values into different classes. Figure 5 illustrates the landscape complexity mapping in test sites in Indonesia. The number of clusters is determined in each site and the value is scaled (on the left) then, the distribution is grouped into three level of landscape complexity following the natural breaks

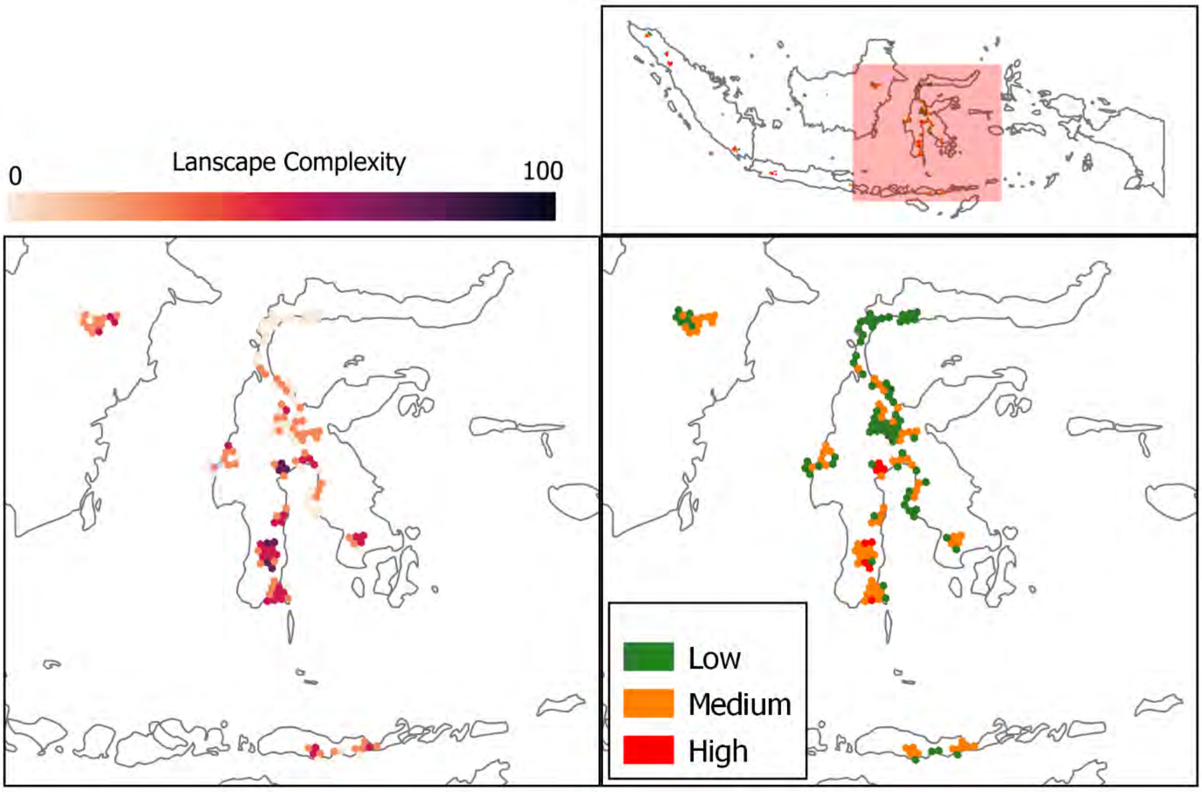

Figure 6: From landscape fragmentation to landscape complexity in Indonesia

6.3.3 Presence of the agrocommodities

A site could only be established as test site for a commodity if it has evidence of the presence of that commodity. To quantify the presence/abundance of the commodity, several data sources were used; Some of the data sources included Trase[2] which provided supply chains and geospatial data for all of the EUDR commodities except forrubber. From this platform, it was possible to extract the production areas and the volume of production in the main producing countries until 2022. Furthermore, the Global Forest Watch data collection[3] was filtered to extract information on industrial and household plantations, oil palm concessions, artisanal permits for annual timber cutting, wood production and forest reserves, just to name a few. Finally, national land cover maps of different countries produced after 2020 were used to extract commodities. In Figure 6, cocoa, rubber and oil palm classes were extracted from the land cover classification of Cote d'Ivoire 2020[4]. The area covered by cocoa, combined with the production were used to estimate the abundance in the sites. For countries with little or no information

[2] https://trase.earth

[3] https://data.globalforestwatch.org/search?collection=Dataset

[4] https://africageoportal.maps.arcgis.com/apps/webappviewer/index.html?id=88c2493e722546c09c2a0a8b394c4454

on the wood exploitation, the forest cover percentage was considered. The data on the abundance of the agrocommodities was scaled and classified into three level of abundance (Low – Medium – High) based on the Jenks natural breaks as described earlier. Figure 6 illustrate the spatial distribution of all the main commodities within the sites in Cote d'Ivoire. It can be seen that there is an overlap between the distribution of cocoa, rubber and wood. We could expect a high landscape complexity in those sites.

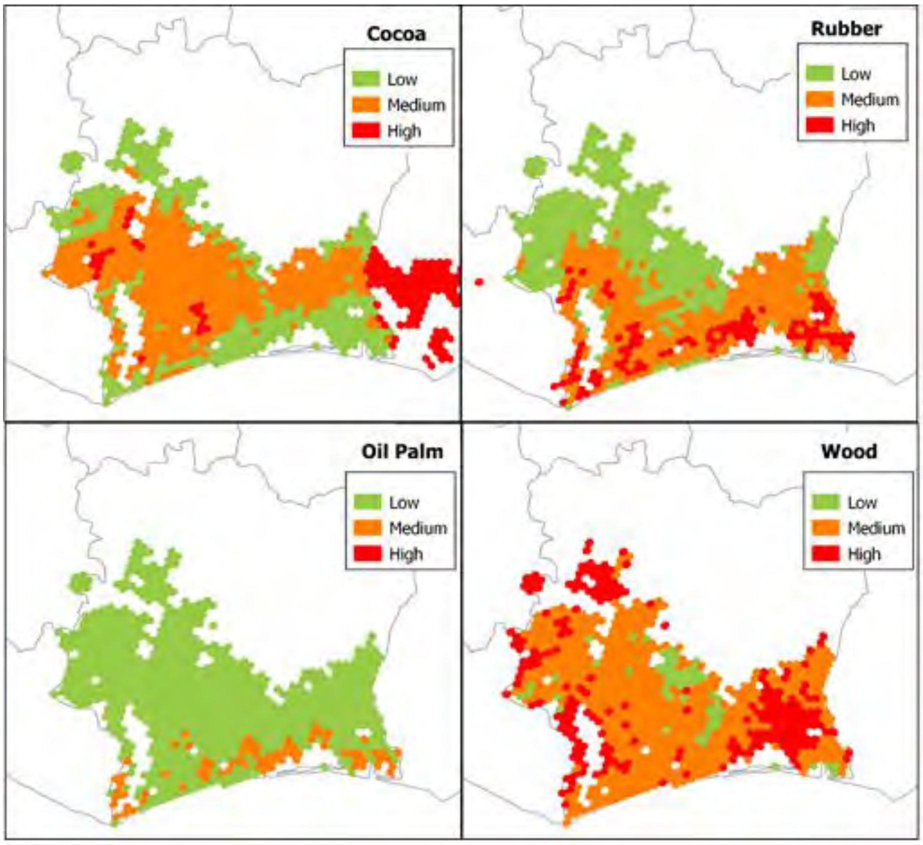

Figure 7: Presence and relative abundance of the EUDR agrocommodities in Cote d'Ivoire

6.4 Test site selection

6.4.1 Cocoa

The number of test sites per location and per country was defined in Table 6. The deforestation risk, the landscape complexity and the abundance of the agrocommodities were combined to identify the potential test sites. Taking the example of cocoa, 10 test sites are distributed among six countries as presented in Table 5. Because of the limited amount of in situ data in Ecuador, only one test site eligible; the number of test sites in Indonesia was then raised to two.

Table 6: Number of test sites per country for cocoa with the availability of in situ data

Countries	Nr test sites	In situ datacounts
Cote d'Ivoire	3	1576
Ecuador	2	1
Ghana	2	282
Nigeria	1	261
Cameroon	1	29
Indonesia	1	100

In Africa, the deforestation risk and the landscape complexity are mostly Medium to High. These characteristics were combined as presented in Table 6, and the selection process followed the steps illustrated in Figure 6. Since the combination of different level of deforestation risk, landscape complexity and presence of cocoa returned multiple sites, samples were selected based on the abundance of in situ data.

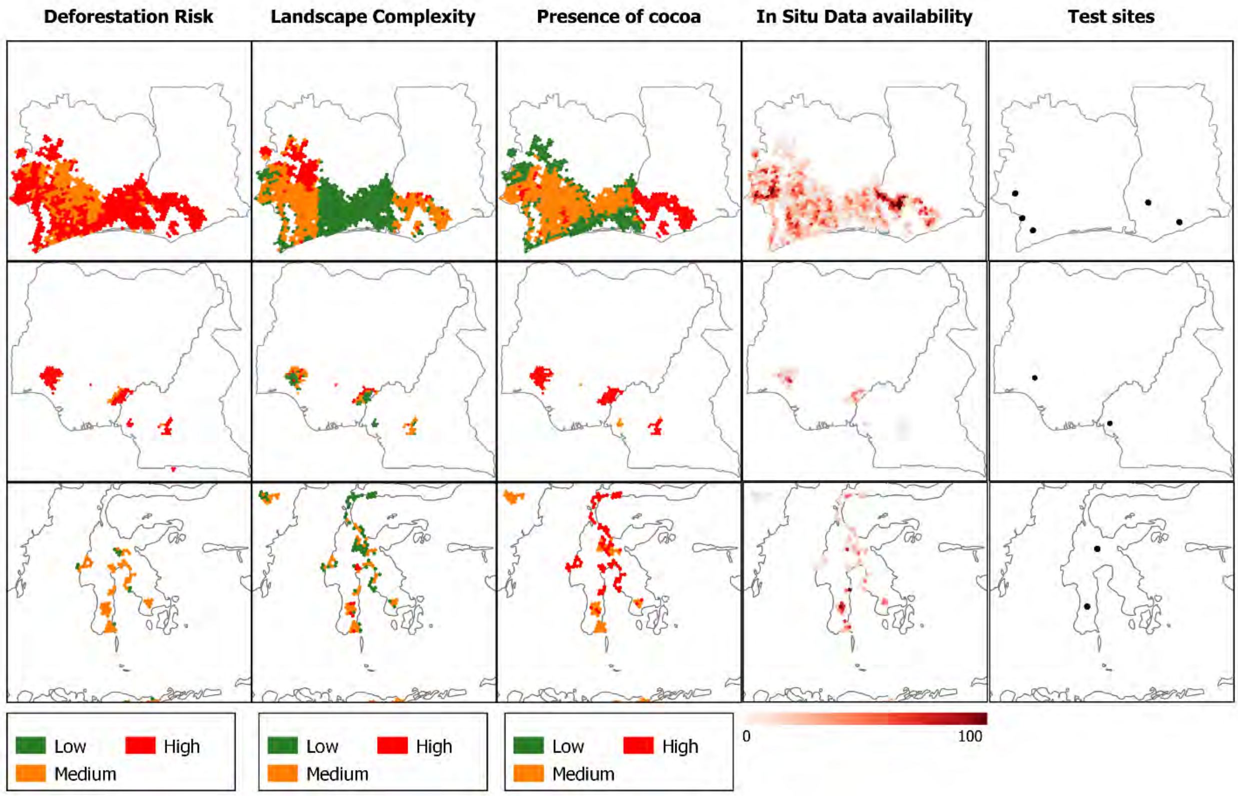

Figure 8: Steps for test sites selection in (from top to bottom) Ghana and Côte d'Ivoire, Nigeria and Cameroon and Sulawesi in Indonesia.

In Cote d'Ivoire, the test sites have a Medium to Low landscape complexity and a High deforestation risk, with a minimum of 438 ha of in situ data coverage. In Ghana, the deforestation risk is High and the landscape complexity is Medium with up to 1307 ha of in situ data. Indonesia was the only country with a combination of Low deforestation risk test sites and Low landscape complexity, corresponding to natural forest areas. In Cameroon and Nigeria, the test sites have a High risk of deforestation and a Low landscape complexity with up to 465 ha of in situ data coverage. Finally in Ecuador, the sample has a High deforestation risk and a Medium landscape complexity, supported by only 7 counts of in situ data for cocoa.

Table 7: Description of the test sites for cocoa

Countries	Deforestation risk	Landscape complexity	Presence of cocoa	In situ data
Cote d'Ivoire (3 test sites)	M	M	H	178 counts (438 ha)
	H	L	M	393 counts (639 ha)
	H	M	M	319 counts (673 ha)
Ghana (2 test sites)	H	M	H	3607 counts (1307 ha)
	H	M	H	643 counts (218 ha)
Indonesia (2 test sites)	L	L	H	12 counts (11 ha)
	M	M	M	3210c counts (967 ha)
Cameroon (1 test site)	H	L	M	268 counts (102 ha)
Nigeria (1 test site)	H	L	M	618 counts (465 ha)
Ecuador (1 test site)	H	M	H	7 counts

6.4.2 Coffee

Test sites for coffee were located in four countries including Brazil, Columbia, Ethiopia and Vietnam (Table 8). The most amount of in situ data is found in Ethiopia with 571 counts corresponding to 400 ha.

Table 8: Number of sample and in situ data in the country of interest

Countries	Nr test sites	In situ data (counts)
Brazil	2	14 (92 ha)
Columbia	3	33 (9 ha)
Ethiopia	3	571 (400 ha)
Vietnam	1	115 (87 ha)

The test sites were evaluated using four parameters as presented in Figure 8. In Ethiopia (top row), the deforestation risk and landscape complexity were Medium and Low respectively, and large amount of data available (over 75%). The site has a high presence of coffee, compared to Brazil were the coffee abundance was relatively low.

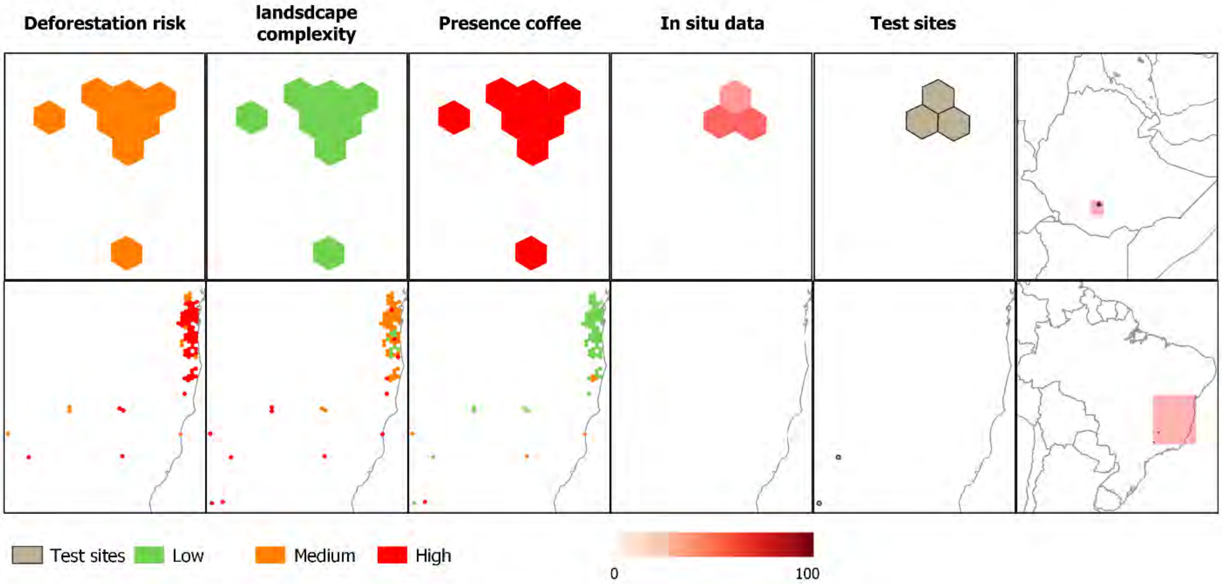

Figure 9: Processing steps for the selection of test sites for coffee in Ethiopia (top) and Brazil (bottom)

Table 9 show a description of the test sites for coffee. The risk of deforestation of the test sites is equally distributed between Medium and High and the landscape complexity is shared between Low and High and only 2 samples in Ethiopia had a Medium landscape complexity.

Table 9: Description of the test sites for coffee

Countries	Deforestation risk	Landscape complexity	Presence of coffee	In situ data (counts)
Ethiopia	M	L	H	122
	M	M	H	293
	M	M	H	278
Vietnam	H	H	H	52
	H	H	H	63
Brazil	M	H	H	9 (70 ha)
	H	H	H	5 (22 ha)
Columbia	M	L	H	13
	H	L	M	18 (8 ha)
	H	L	M	2 (1ha)

6.4.3 Soy

The selection of test site selection for soy was carried in three countries including Argentina, Brazil and the USA (Figure 9). The criteria of selection included the deforestation risk, the landscape complexity and the amount of in situ data available.

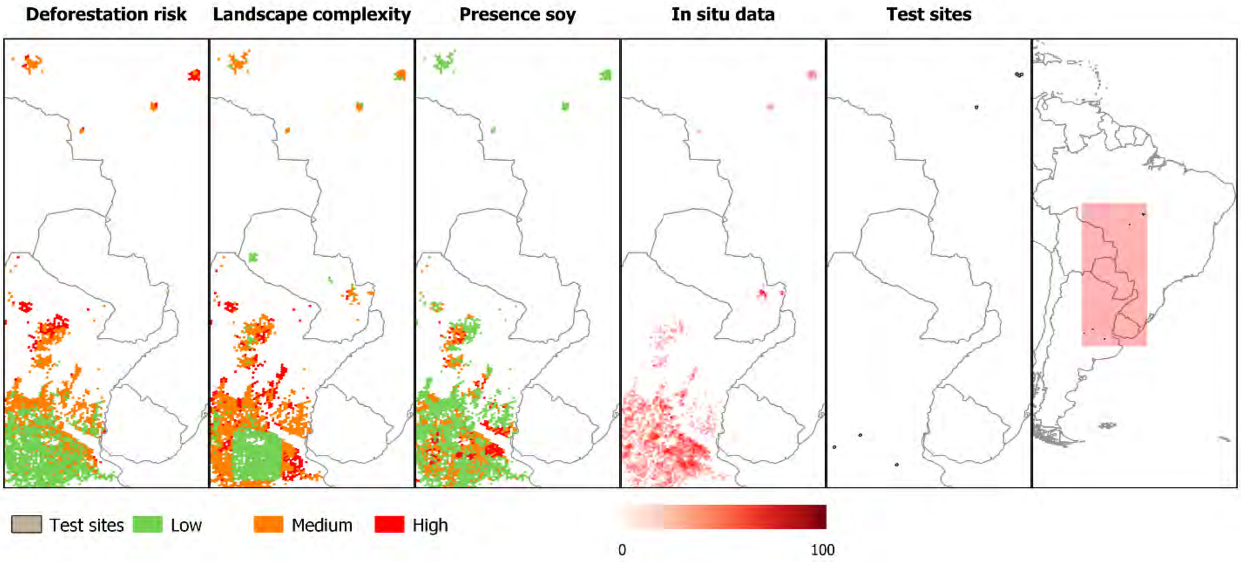

Figure 10: Processing steps for the selection of test sites for soy in South America

A description of the test sites, summarized in Table 10 illustrates the availability of in situ data for each of the test site. The largest amount of in situ data was found in Argentina, with a relative Medium presence of soy plantations within a wide variety of landscape complexity associate with a Low to High level of deforestation risk. In the US on the other hand, the deforestation risk was Low in the test site due to a relatively low percentage of forest cover, with a Medium landscape complexity.

Table 10: Description of the test sites for soy

Countries	Deforestation risk	Landscape complexity	Presence of soy	In situ data (counts)
Brazil	M	M	L	7
	M	M	L	10
	H	M	L	9
	H	M	L	8
Argentina	L	M	M	34 (4306 ha)
	M	L	M	32 (1251 ha)
	M	H	H	41 (1682 ha)
USA	L	M	H	12
	L	M	H	12
	L	M	H	11

6.4.4 Oil palm

The final selection of test sites for oil palm were distributed across three countries including Cote d'Ivoire, Indonesia and Thailand. Because of the absence of site with oil palm plantations in Liberia, Vietnam and Malaysia, no test sites were selected in those countries. The selection of the test sites was based on the integration of the deforestation risk, landscape complexity and availability of in situ data in each site (Figure 10). The deforestation risk was higher in Cote d'Ivoire compared to Indonesia and the presence of oil palm in test sites was relatively low.

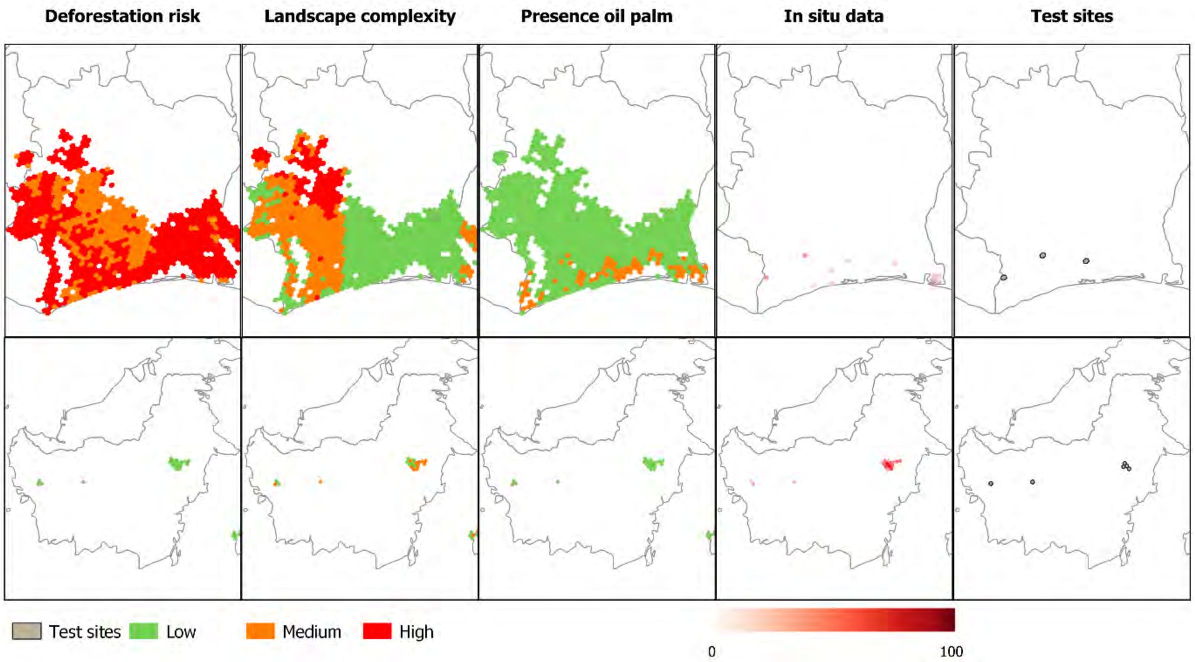

Figure 11: Processing steps for the selection of test sites for oil palm in Côte d'Ivoire (top) and Indonesia (bottom)

The distribution of test sites for oil palm was redistributed between three countries namely Cote d'Ivoire, Indonesia and Thailand because of a lack of lack of in situ data in certain countries. The description of the test sites is summarized in Table 11. All of the test sites have a very high risk of deforestation even if the presence of oil palm plantations is not abundant, suggesting the presence of other crops mainly represented by rubber and cocoa. Most available data were gathered in Indonesia with up to 850 counts in a single site.

Table 11: Description of the test sites for oil palm

Countries	Deforestation risk	Landscape complexity	Presence of oil palm	In situ data (counts)
Cote d'Ivoire	H	M	H	43
	H	L	M	35
	M	L	M	19
Indonesia	H	L	L	24
	H	L	L	38
	H	M	L	850
	H	M	L	307
	H	M	L	384
	H	L	L	685
Thailand	H	M	M	20

6.4.5 Rubber

Based on Table 4, test sites for rubber were distributed in Africa (Cote d'Ivoire and Liberia) and in Asia (Vietnam, Indonesia, Malaysia and Thailand). The selection was based on the deforestation risk, the landscape complexity and the availability of in situ data (Figure 11).

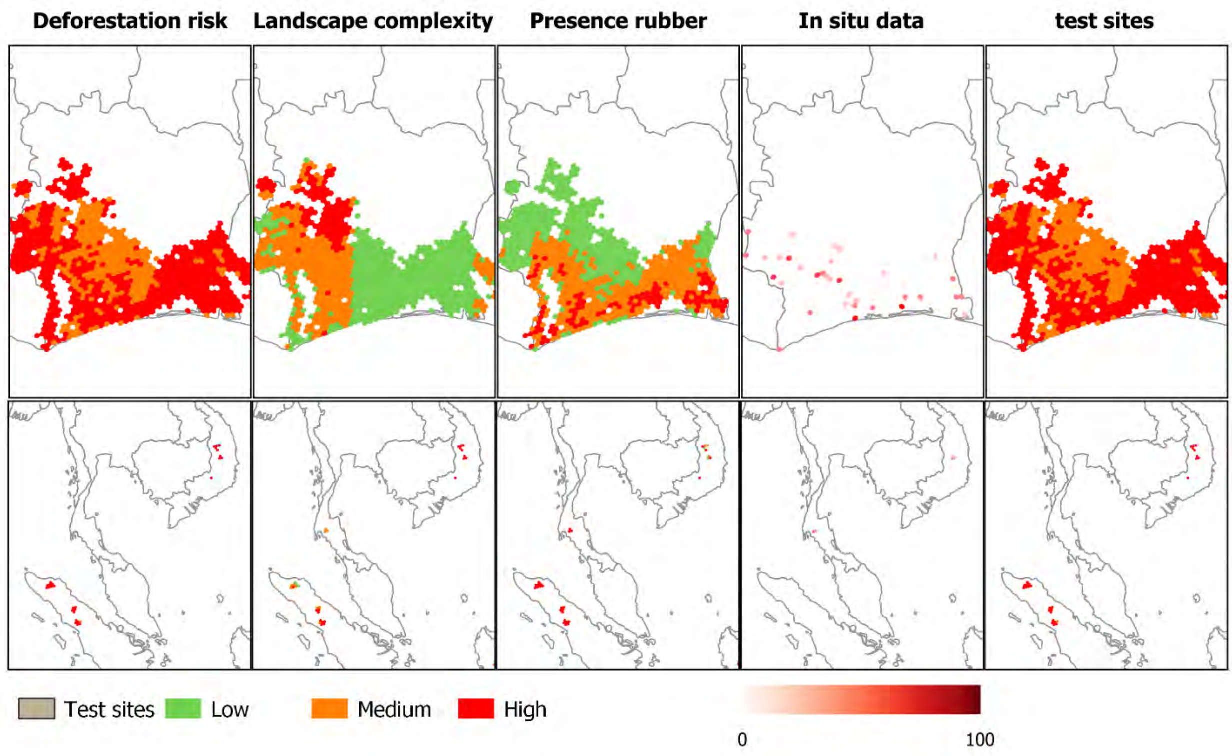

Figure 12: Processing steps for the selection of test sites for rubber in Côte d'Ivoire (top) and Thailand (bottom)

Test sites for rubber were gathered in four countries: Cote d'Ivoire, Gabon, Thailand and Vietnam. The test sites area mostly associated with a high risk of deforestation and showed a relatively low cocoa abundance in Africa compared to Asian countries. This could be explained by the presence of other commodity such as cocoa, oil palm or wood. The landscape complexity is relatively Medium to Low suggesting large plantations instead of fragmented small-scale farming.

Table 12: Description of the test sites for rubber

Countries	Deforestation risk	Landscape complexity	Presence of rubber	In situ data (counts)
Cote d'Ivoire	M	M	M	20
	M	M	M	14
	H	L	L	6
	H	L	H	48
	H	L	M	22
Gabon	H	M	M	8
Vietnam	H	H	L	7
	H	H	H	7
Thailand	H	M	H	11
	H	M	H	18

6.4.6 Wood / timber

The selection of test sites for wood was based on two steps including the level of deforestation risk and the landscape complexity. Explicit data on the availability of in situ data were missing in the site, therefore the expert judgement was used for the final selection of the test sites (Figure 12).

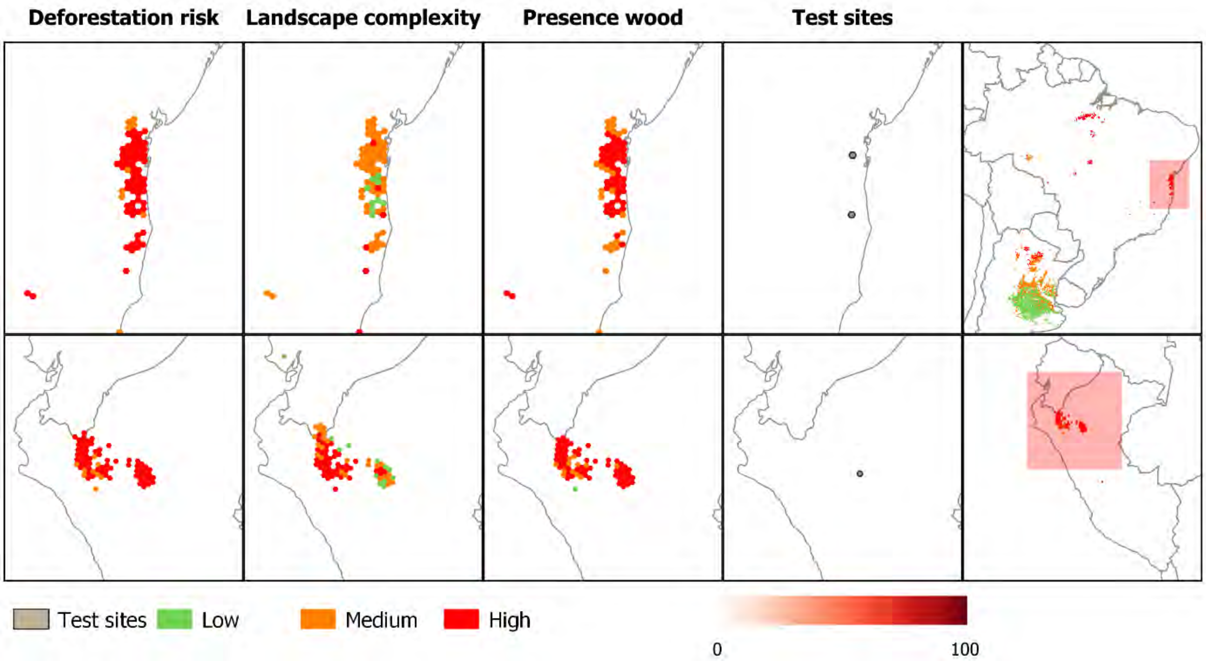

Figure 13: Processing steps for the selection of test sites for timber

For timber, the test sites were distributed between Brazil, Cote d'Ivoire, Gabon, Indonesia, Peru, the USA and Vietnam as presented in Table 13. The presence of timber was extracted from data on the percentage of tree cover combined with information on forest concession where available (mostly in Africa) to distinguish between planned and unplanned logging activity. The test sites both have a High risk of deforestation and a high abundance of timber. Concerning the landscape complexity, the level of fragmentation in Low to Medium except in Peru and Vietnam where there is a High level of fragmentation. A Low to Medium landscape complexity suggest the proximity to natural forest.

Table 13: Description of the test sites for wood

Countries	Deforestation risk	Landscape complexity	Presence of wood
Brazil	H	L	H
	H	M	H
Peru	H	H	H
USA	M	M	M
Cote d'Ivoire	H	L	H
Gabon	H	L	H
	H	L	H
Vietnam	H	H	H
Indonesia	H	M	H

6.4.7 Cattle/beef

Test sites for cattle were located in South and North America, in Brazil, Argentina and the USA. Figure 13 illustrates the steps of the selection of test sites in Brazil and Argentina. It was based on the integration of information related to the level of deforestation risk, landscape complexity and availability of in situ data.

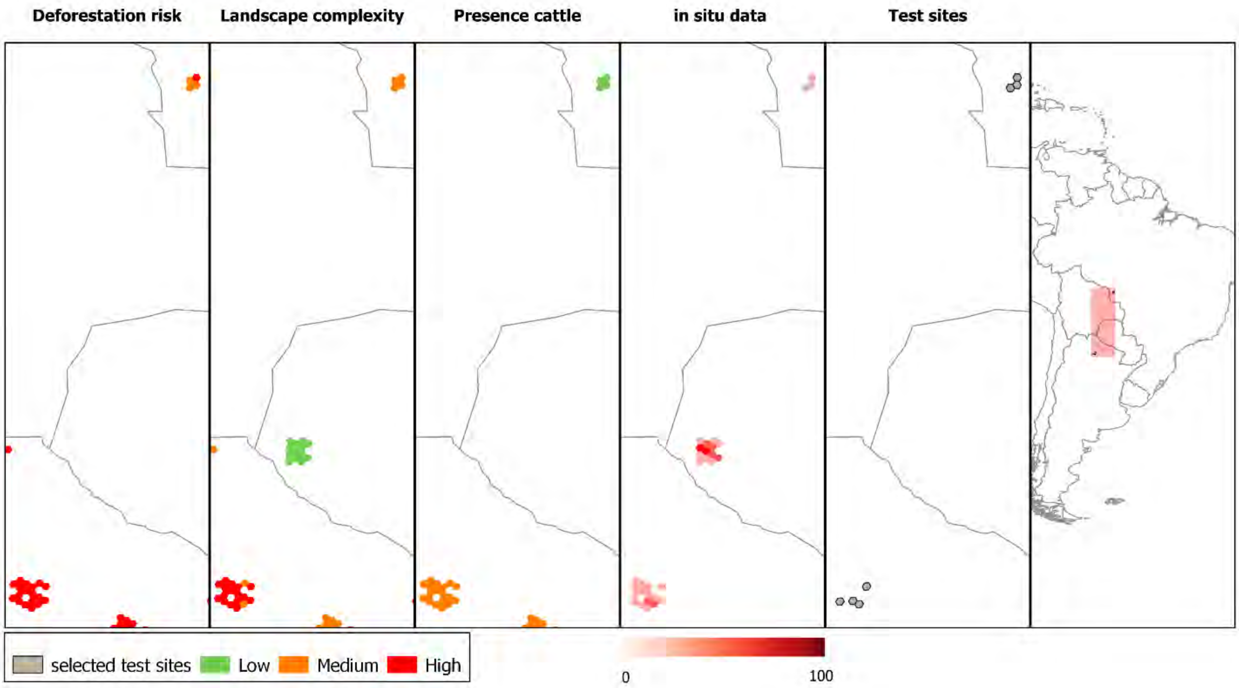

Figure 14: Processing steps for the selection of test sites for cattle in South America

The final test sites for the cattle were to selected in Brazil, Argentina and the USA (Table 14). The data for cattle were derived from data on grassland/pasture and information about beef product exportation from Trase. It was found that cattle production is associated with Medium to High risk of deforestation, and their presence was Low to Medium in the site suggesting the presence of other land management practices. Cattle also occurs in a Medium to High landscape complexity, further supporting the presence of other land management practices in the sites.

Table 14: Description of the test sites for cattle

Countries	Deforestation risk	Landscape complexity	Presence of cattle	In situ data counts
Brazil	M	M	L	2
	M	M	L	3
	H	M	L	2
Argentina	H	H	M	4
	H	H	M	7
	H	M	M	5
	H	H	M	2
USA	M	M	M	
	M	M	M	
	M	M	M	

7 Conclusion

As part of this requirement baseline assessment a policy review focussing on the EUDR was conducted followed by a summary of the User Requirements emanating from the extensive consultation with CNAs through guided interviews a first living lab and an analysis of the potential risks associated with some of the outlined requirements. The latest developments were reviewed related to relevant technical solutions currently available to meet the different user requirements, as well as the gaps in the current methodologies which then led to a description of a proposed technical approach for the current project. This was then followed with a structured approach on how the test and demonstration sites were selected for the different commodities.

The key gaps identified as par the review of the state-of-the-art are primarily linked to the following main topic:

- Single crop focus as opposed to tackling multiple commodities.
- Link to deforestation not always assessed or explicit.
- Limited focus on commodity expansion over several years.
- Smallholder production systems (e.g. Agroforestry) are not always properly characterised as they cover small areas, are difficult to map and can often be confused with forest areas.

Some key methodological challenges were also outlined:

- Data accessibility and pre-processing to ensure analysis ready data with constant quality are available.
- State-of-art ML/DL models require large amount of representative datasets to be applicable globally and their integration in OpenEO workflows also represents a substantial challenge.
- While the envisaged Open architecture of the WAC prototype system can integrate private data into its workflows, special attention must be given to addressing the privacy requirements of the CNAs.

The detailed feedback received from CNAs made it possible to outlined the requirements for the WAC protype system following a two-step approach:

3. A rapid potential non-compliance detection system followed by
4. An evidence-based verification system.

The requirements expressed were translated in a set of functional and technical characteristics that will be tested in the next step focusing on the benchmarking of methods with the aim to assess the feasibility and limitations of addressing some of these requirements. The final specifications for the WAC prototype system will be documented in the Algorithm Theoretical Baseline Document (ATBD) and Technical Specifications (TS) and will be implemented over the selected demonstration sites ensuring CNAs and other stakeholders are consulted at each key step during the process.

8 References

Abu, I. O., Szantoi, Z., Brink, A., Robuchon, M., & Thiel, M. (2021). Detecting cocoa plantations in Côte d'Ivoire and Ghana and their implications on protected areas. Ecological indicators, 129, 107863.

Bardach, E., & Patashnik, E. M. (2023). A practical guide for policy analysis: The eightfold path to more effective problem solving. CQ press.

Bourgoin, C., Ameztoy, I., Verhegghen, A., Desclée, B., Carboni, S., Bastin, J., ... & Achard, F. (2024). Mapping global forest cover of the year 2020 to support the EU regulation on deforestation-free supply chains.

Broekx, R (2023). How to create a pixel-based crop classification model. Techniques and Architectures for Remote Sensing and Agricultural Applications. https://medium.com/superlinear-eu-blog/how-to-create-a-pixel-based-crop-classification-model-88f8f59fbff3

Descals, A., Gaveau, D. L. A., Wich, S., Szantoi, Z., and Meijaard, E. (2024). Global mapping of oil palm planting year from 1990 to 2021, Earth Syst. Sci. Data, 16, 5111–5129, https://doi.org/10.5194/essd-16-5111-2024.

Dong, R., Li, W., Fu, H., Gan, L., Yu, L., Zheng, J., & Xia, M. (2020). Oil palm plantation mapping from high-resolution remote sensing images using deep learning. International Journal of Remote Sensing, 41(5), 2022-2046.

Hansen M. C., Potapov P.V., Moore R., Hancher M., Turubanova S.A., Tyukavina A., Thau D., Stehman S.V., Goethz J. […] and Townshend J.R.G. (2013). High resolution global maps of 21st centuray forest cover. Science vol 342, issue 6160. Pp. 850 – 853 https://www.science.org/doi/10.1126/science.1244693

Jenks, G. F. 1967. The Data Model Concept in Statistical Mapping, International Yearbook of Cartography 7: 186–190.

Kalischek, N., Lang, N., Renier, C. et al. Cocoa plantations are associated with deforestation in Côte d'Ivoire and Ghana. Nat Food 4, 384–393 (2023). https://doi.org/10.1038/s43016-023-00751-8

Li, H., Song, X-P., Hansen, M.C. et al., (2023). Development of a 10-m resolution maize and soybean map over China: Matching satellite-based crop classification with sample-based area estimation. Remote Sensing of Environment, 294, 113623, https://doi.org/10.1016/j.rse.2023.113623

Maskell, G., Chemura, A., Nguyen, H., Gornott, C., & Mondal, P. (2021). Integration of Sentinel optical and radar data for mapping smallholder coffee production systems in Vietnam. Remote Sensing of Environment, 266, 112709.

Masolele, R. N., De Sy, V., Marcos, D., Verbesselt, J., Gieseke, F., Mulatu, K. A., ... & Herold, M. (2022). Using high-resolution imagery and deep learning to classify land-use following deforestation: a case study in Ethiopia. GIScience & Remote Sensing, 59(1), 1446-1472.

Masolele, R. N., Marcos, D., De Sy, V., Abu, I. O., Verbesselt, J., Reiche, J., & Herold, M. (2024). Mapping the diversity of land uses following deforestation across Africa. Scientific Reports, 14(1), 1681.

Neeff, T., Merle, C., D'Annunzio, R., O'Brien, V., Gill, G., Poultouchidou, A. and Fox, J. 2024. Forest monitoring for transparent commodity supply chains. Rome, FAO. https://doi.org/10.4060/cd3280en

Parente, L., Sloat, L., Mesquita, V., Consoli, D., Stanimirova, R., Hengl, T., ... & Stolle, F. (2024). Annual 30-m maps of global grassland class and extent (2000–2022) based on spatiotemporal Machine Learning. Scientific data, 11(1), 1-22.

Somching, N., Wongsai, S., Wongsai, N., & Koedsin, W. (2020). Using machine learning algorithm and landsat time series to identify establishment year of para rubber plantations: A case study in Thalang district, Phuket Island, Thailand. International Journal of Remote Sensing, 41(23), 9075-9100.

Song, X. P., Hansen, M. C., Potapov, P., Adusei, B., Pickering, J., Adami, M., ... & Tyukavina, A. (2021). Massive soybean expansion in South America since 2000 and implications for conservation. Nature sustainability, 4(9), 784-792.

Van Tricht, K., Degerickx, J., Gilliams, S. et al. (2023). WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping. Earth System Science Data, 15, 12, 5491-5515. https://doi.org/10.5194/essd-15-5491-2023

Ribeiro, V., Mendes, F. D. S., Singh, C., & Heilmayr, R. (2024). Leveraging remote sensing for transparency and accountability in Amazonian commodity supply chains. One Earth, 7(11), 1935-1940.

Université Catolique de Louvain (2020). Building a crop map for South Africa using Sen2-Agri. Knowledge Package at GeoKnowledgeHub. https://gkhub.earthobservations.org/packages/ajbv5-e5b80

VITO Remote Sensing (2024). Launch of the WorldCereal crop mapping system. News article. https://esa-worldcereal.org/en/launch-worldcereal-crop-mapping-system

Wang, Y., Hollingsworth, P. M., Zhai, D., West, C. D., Green, J. M., Chen, H., ... & Ahrends, A. (2023). High-resolution maps show that rubber causes substantial deforestation. Nature, 623(7986), 340-346.

This page is intentionally left blank!

Annex 1: Guided Interview Questions

BLOCK A: General understanding of EO data usage:

1) Could you provide examples and/or an overview on how Earth Observation data can support the EUDR implementation (i.e. for compliance monitoring purposes)?

2) What kind of EO-based information and tools do you envisage to use?

BLOCK B: Detailed questions to ESA WAC and EUDR Information Requirements for Due Diligence

Questions specific to ESA WAC: related to risk assessment of crop expansion (part of EUDR Art 29)

3) What are the commodities and related countries for you to consider as a priority to develop and test the use of EO?

4) In the absence of specific priorities, could you outline criteria to consider to identify suitable sites to develop and test our approach such as commodity production trend, overall production, and distance to forests?

Questions based on EUDR Information Requirements for Due Diligence:

Geolocation of all plots of land where the commodities are produced:

(from the regulation: geolocation (or polygons for plots > 4ha for all commodities other than cattle) of all plots of land where the products were produced (or cattle has been kept))

5) Do you need data/methods for plausibility checks on geolocation and areas of the reported commodity?

6) Are you familiar with the requirements of the EU Information System for EUDR; for example, that the geospatial data should be provided in GeoJSON format?

7) Do you need data/methods for verification of occurrence or not of deforestation (i.e. change in land use from Forest to agricultural use) on 31.12.2020 and thereafter?

(Related to evidence that products have been produced without deforestation after 31 December 2020: comparing forest cover at the point of production with forest cover after 31 December 2020.)

(Some aspects for the deforestation assessments: Be able to distinguish between tree cover loss (e.g. removal of dead or diseased trees), which is acceptable, and conversion of forest to agricultural use, which is not. Be able to distinguish between 'forest' (see definition in EUDR) and other less heavily wooded land.)

8) Do you need data/methods for verification of occurrence or not of degradation on 31.12.2020 and thereafter?

(For wood products, they must also be free of forest degradation after 31 December 2020. 'Forest degradation' means structural changes to forest cover, taking the form of the conversion of: (a) primary forests or naturally regenerating forests into plantation forests or into other wooded land; or (b) primary forests into planted forests (Article 2(7)).

9) Do you need any ancillary data such as Protected Areas/Land Use Land Cover maps? Is there any additional supporting spatial data sources that would need to be considered as part of the control mechanisms?

Examples: protected areas, land Use/land tenure (e.g.; Forest / Agricultural plantations…)

10) Resolutions: which frequency of updates do you think is needed (yearly, monthly, NRT) and what is in your opinion the optimal spatial resolution for deforestation / agricultural mapping?

<u>BLOCK C</u>: Challenges and next steps

11) Commodity-specific challenges: Given the diversity of commodities covered by the EUDR, are there specific commodities or product categories that you anticipate posing greater challenges for compliance monitoring and enforcement, and why?

12) Resource Constraints: Are there resource constraints that limit your ability to effectively enforce the EUDR, and how could satellite data or tools help to address these constraints?

13) Are there any additional aspects not covered above that we should consider?

Annex 1: ESA WAC LL 1 Workshop Report

See Separate documents delivered on 20/11/2024 as listed below:

Pos.	Description	Project Server Link	Amount
1	Living Lab #1 Presentations	https://uranus.gaf.de/index.php/s/D7RitzZLxaX6EDQ	5
2	Living Lab #1 minutes	https://uranus.gaf.de/index.php/s/FpaAZB5iPPCes4s	1

End of Document