An Earth Observation Data Cube System for Monitoring Deforestation and Urban Development Trends in Kenya: A Case Study of Murang’a County

by

Kanda Kenneth Kirop

Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Information Technology (MSc IT) at Strathmore University

School of Computing and Engineering Sciences, Strathmore University, Nairobi. Kenya

March, 2025

Declaration

This research proposal is my original work and has not been presented for a degree in any other university. No part of this proposal may be reproduced without the prior permission of the author and/or Strathmore University. All other sources of information cited herein have been duly acknowledged.

Signature .... Date .....**27th March 2025**.................

Name: Kanda Kenneth Kirop

Student Number: 153116

SUPERVISOR'S DECLARATION

This research proposal has been submitted for review with my approval as a university supervisor.

Signature……… ……………Date………27/03/2025**………**…………

Professor Ismail Ateya

School of Computing and Engineering Sciences,
Strathmore University.

Table of Contents

Abstract

This research considers the development of an earth observation (EO) data cube system to facilitate integration of spatial data analysis in decision making in the country by using Murang'a county as a case study. The primary goal is to develop an infrastructure that provides for effective and efficient analysis of open-access EO data so as to monitor the state of natural resources such as forests and to also track the urban development trends. This infrastructure should have the capacity to detect trends in order for decisions to be made on events such as deforestation and uncontrolled urban developments. Technically, the data cube system aims to overcome challenges in accessing, analyzing and processing huge amounts of geospatial datasets. This is to be achieved through developing the necessary data pipelines to acquire data from open-access satellite image providers such as Sentinel satellites. This data is to be converted from the raw EO data to more usable formats. The case study will be on Murang'a County due to its forests as well as the continued urban development. The forest cover and the urban development patterns will be analyzed from the data and the time-series analysis is expected to provide insights on how to address the negative effects of such changes. The research proposes to use open-source tools and technologies to implement this system with the foundation being on the Open Data Cube framework.

Keywords: analysis-ready data, data cube, earth observation data, environmental monitoring, floods, geospatial, spatial planning

Acknowledgments

I am extremely grateful to Almighty God for His riches in grace and mercies in granting life, faith in Him and His Providence to be able to undertake my postgraduate studies. I would also like to express my heartfelt gratitude to my supervisor, Prof. Ismail Ateya, for his invaluable guidance and support throughout my research. His insightful feedback and expertise have significantly shaped this thesis and enhanced my understanding of analysis of earth observation data.

I also wish to acknowledge the support of the Master of Science in Information Technology class of 2025 for fostering an intellectually stimulating environment that is necessary to develop quality work. Their engaging questions and thought-provoking discussions have inspired me to delve deeper into my research topic. Lastly, I would like to extend my deepest appreciation to my wife, family and friends for their unconditional love and encouragement throughout this journey.

Abbreviations/Acronyms

API	Application Programming Interface
ARD	Analysis Ready Data
AWS	Amazon Web Services
BDC	Brazilian Data Cube
CDC	China Data Cube
CEOS	Committee on Earth Observation Satellites
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CPU	Central Processing Unit
DGGS	Discrete Global Grid System
EO	Earth Observation
EODC	Earth Observation Data Cube
ESA	European Space Agency
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GPU	Graphics Processing Unit
IWMI	International Water Management Institute
KDC	Kenyan Data Cube
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LaSRC	Land Surface Reflectance Code
LiMES	Live Monitoring of Earth Surface
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
ODC	Open Data Cube
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
SAR	Synthetic Aperture Radar
SDC	Swiss Data Cube
STL	Seasonal-Trend decomposition using LOESS
SVM	Support Vector Machine
USGS	United States Geological Survey
WOfS	Water Observations from Space

Definition of Terms

Analysis Ready Data - Earth Observation data that has been standardized to a specific baseline, to a level sufficient for immediate analysis with little extra work from the user. This entails things such as radiometric calibration, geometric correction, and atmospheric correction. (Lewis et al., 2019).

Atmospheric Correction - It is the calculation that removes the scattering or absorbing effects of the atmosphere on the satellite image so that an accurate surface reflectance value is obtained for true quantitative analysis. (Chatenoux et al., 2021).

Data Cube - Earth Observation data is indexed along the space, time, and data type axes within a multi dimensional array to permit structured query and spatial and temporal analysis. (Chatenoux et al., 2021).

Discrete Global Grid System - Multi resolution equal area partitioning of the earth’s surface into cells enables global data to be integrated and analyzed at different spatial scales. (Wang et al., 2018).

Earth Observation (EO) - The acquisition and analysis of the environmental and geophysical features of the earth from space using sensors mounted on satellites and aircraft to observe the global systems. (Gomes et al., 2021).

Earth Observation Data - Data collected through remote sensing include spectral reflectance, temperature, elevation, and many other Earth's physical properties that can be examined and analyzed. (Gomes et al., 2021).

Flood Risk Management - This is a proactive approach towards reducing the consequences of flooding through an integrated assessment, action, and structure strategy. (International Water Management Institute, 2024).

Geographic Information System - An integrated digital system enabling capturing, storing, analysing, and visually examining location-based data and its attributes. (Li et al., 2016).

Geospatial Data - Data that has an association point on the earth, such as objects, events or phenomena, in conjunction with their spatial coordinates and descriptive attributes. (Hu et al., 2018).

Google Earth Engine - A cloud-based Google geospatial processing platform capable of planetary scale analysis of environmental datasets, thanks to Google's computational abilities. (Amani et al., 2020).

Land Cover - A physical geographical area that includes vegetation, water, rocks, bare soil, man-made structures, and snow, which can be seen directly, or through remote sensing. (Ferreira et al., 2020).

Land Use - Human activities and economic activities focused on a particular area of the land, for example, residential, agricultural, industrial, conservation or recreational. (Ferreira et al., 2020).

Machine Learning - Computational approaches that enable systems to identify patterns, develop insights and/or make decisions from data through training and experience rather than instructed programming. (VoPham et al., 2018).

Open Data Cube (ODC) - Open-source framework for managing, processing, and analysing Earth observation data through spatially and temporally aligned data structures. (Killough, 2019).

Remote Sensing - Science of acquiring information about Earth's surface through detecting and measuring radiation reflected or emitted from objects, typically using airborne or satellite-based sensors. (Ma et al., 2020).

Satellite Imagery - Digital images of Earth captured by using satellites by various sensors, including optical, radar, and multispectral instruments, providing regular global coverage. (Gorelick et al., 2017).

Spatial Planning - Deliberate process of apportioning and managing spatial resources to enhance land use, infrastructure development, and environmental protection across different scales. (Makueni County Government, 2021).

Spatial Resolution - The smallest distinct detail in an image, measured as the ground sampling distance characterised by each pixel, determining the level of evident detail in Earth Observation data. (Dwyer et al., 2018).

Sustainable Development Goals - United Nations' framework of seventeen (17) interconnected objectives considered to achieve global environmental sustainability, social equality, and economic development by 2030. (United Nations, 2015).

Time Series Analysis - A statistical approach for the analysis of sequences of data points over time aiming at extracting meaningful patterns, trends and seasonal variations. (Malinowski et al., 2020).

Wetland Inventory Mapping - An alternative methodology that involves the systematic identification and demarcation of wetland ecosystems using remote sensing and GIS techniques geared towards monitoring these areas as well as managing them effectively. (Amani et al., 2020).

List of Figures

List of Tables

List of Equations

Equation Number	Equation Description	Formula
1	Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{NIR - RED}{NIR + RED}$
2	Enhanced Vegetation Index (EVI)	$EVI = G \times \frac{NIR - RED}{NIR + C_1 \times RED - C_2 \times BLUE + L}$ where $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$, $L = 1$
3	Normalized Difference Water Index (NDWI)	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
4	Normalized Difference Built-up Index (NDBI)	$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$

Chapter 1: Introduction

1.1 Background

As reported by the Ministry of National Treasury and Economic Planning in 2024, the Government of Kenya has embarked on a serious tree planting initiative. The main goal of this initiative is to combat and mitigate climate change effects. Particularly, the efforts aim to reduce greenhouse gases, reverse trends in deforestation and stabilize climate patterns. Switching to land-use in the country, the Constitution of Kenya (2010) provides a legal framework which requires county governments to develop County Spatial Plans in order to ensure effective and efficient use of land resources. This is a requirement that only a few counties have been able to manage and thus, a lot of unplanned developments are undertaken at county level leading to strain of resources or situations where offering public services is strained. These two areas of forest cover and spatial planning are areas that could benefit greatly from the potential use of EO data to support the decision-making.

The Committee on Earth Observation Satellites (CEOS, 2017) developed an architecture to aid in the organization of EO data called the Open Data Cube (ODC) architecture. This framework reduces the complexity of processing large EO datasets by providing indexing and stacking of satellite images in an efficient manner. The goal of the Committee on Earth Observation satellites was to facilitate more informed decision making and policy formulation for the purposes of the greater global conservation of the environment. The ODC architecture lowers the barrier of access to EO data by reducing the technical skill required to access EO data. Gomes et al. (2021) emphasizes the value of freely available remotely sensed data in addressing environmental challenges such as deforestation and natural disasters. Through analysis of such data using the ODC framework, it is possible to gain useful insights into land cover changes, deforestation patters and areas affected by disasters thus becoming a critical tool for decision-makers and policy makers.

In view of the foregoing, it is then important to explore ways of utilizing the ODC technology in supporting the conservation of the environment. Developing a cost-effective platform which combines the available open-access data is an initiative that can augment the current structures in place for making decisions in conserving the environment as well as plan for infrastructure development across the country. Chatenoux et al. (2021) highlights the potential of the data cube architecture in EO data processing through using it to generate Analysis Ready Data (ARD). The provision of analysis ready data quickens the process that analysts undergo in developing analysis products for sharing with the respective decision-makers and policy makers. An example would be a Geographic Information System (GIS) unit in a county government's Lands Department would require such data to conduct their own analysis of utilities in their county. Therefore, a system that collects EO data and provides it in an easily accessible manner is a useful system.

This research takes a case study of Murang'a County which covers around 2,559 square kilometres in central Kenya and has diverse topographical areas (Murang'a County Government, 2018). The county's forests are mainly found in the Aberdare ranges and they face the pressure of increased land coming under farming as well as urban development. As a response to global changes and the local changes, the Murang'a County Climate Change Action Plan (2022-2027) was developed with the goal of growing the forest cover to 10% of the total land area so as to strengthen the county climate resilience and preserve biodiversity (Murang'a County Government, 2023). A comprehensive EO data system would be useful to provide necessary support in monitoring initiatives coming from the action plan thus monitor forestation and deforestation as well as guide evidence-based land use policies.

1.2 Problem Statement

Counties in Kenya such as Murang'a County face substantial challenges in accessing and utilizing spatial data across its key sectors, especially in urban planning, disaster response, and natural resource management. While stakeholders recognize integration of spatial data as essential for addressing unintended urban expansion and environmental degradation, substantial implementation barriers persist (Makueni County Government, 2021). The Murang'a County Climate Change Action Plan (2022–2027) highlights the imperative for data-driven approaches to manage climate-related threats, including floods, landslides, and deforestation, while recognising substantial gaps in spatial data availability and application (Murang'a County Government, 2023). The requirement for accurate spatial data becomes especially severe in disaster response situations, especially within flood-prone areas where detailed terrain and land cover information directly influences mitigation strategy effectiveness (Giuliani et al., 2022). Existing challenges of fragmented data sources and delayed access substantially compromise disaster readiness efforts, resulting in increased vulnerability and increased losses during extreme events (Ferreira et al., 2020). Furthermore, the exponential increase in geospatial raster data volume presents technical difficulties for data management and processing, affecting the ideal utilization of this key information in emergency response and planning applications (Hu et al., 2018).

1.3 Aim / General Objective

The main goal of this research is develop and implement an earth observation data cube system for Murang’a county to provide easy access to analysis ready data to support environmental monitoring and urban planning. The data cube system is to serve as a comprehensive platform for a variety of spatial data analysis with particular focus on forest cover monitoring and urban development trend analysis. This would support the strategic plans of the county with a technological solution that appropriately fits the task at hand. Upon the lessons learnt from the case study, then this can be replicated across counties in Kenya to provide decision-making tools for county governments.

1.4 Research Objectives

i. To review existing earth observation data cube systems.
ii. To identify challenges in the existing earth observation data cube systems.
iii. To design a custom EO data cube system to handle Murang’a county’s data.
iv. To develop the EO data cube system to collect and present analysis-ready data for Murang’a.
v. To test the EO data cube system by developing forest cover and urban development trend analysis maps for Murang’a.

1.5 Research Questions

i. What are the existing earth observation data cube systems?
ii. What are the challenges in the existing earth observation data cube systems?
iii. How to design a custom EO data cube system to handle Murang’a county’s data?
iv. How to develop the data cube system to collect and present analysis-ready data for Murang’a county?
v. How to test the EO data cube system by developing forest cover and urban trend analysis maps for Murang’a?

1.6 Justification

Though there have been substantial efforts to improve spatial data accessibility across important economic sectors such as spatial planning, natural resource management and disaster response, there are significant challenges still in existence. This research proposes a solution to these challenges by developing a platform that utilizes available open-access data and open-source technologies to support the making of decisions and policies to address these key issues. The solution aims to take advantage of a massive amount of EO data that is in existence from the satellites in orbit. In addition to that, research indicates that spatial data is a key determinant in the decisions made for purposes of conserving the environment and executing urban plans. This is well illustrated in the Makueni County Spatial Planning Policy (2021) which points out how improved access to spatial data enhanced the land-use planning and disaster preparedness..

The expected beneficiaries of this research are stakeholders such as county governments and government agencies that are responsible for managing the environment, spatial planning and management of disasters as well as private sector players who seek to leverage EO data for a variety of applications. This spatial data infrastructure is also aimed at supporting existing national infrastructures that are used in providing requisite technical input to inform policy as well as remedial actions. The outcomes are expected to showcase the potential of EO data cubes in analysis of spatial data and capacity to support making decisions and therefore build a case for development of robust systems to provide data availability as well as analytical capabilities.

1.7 Scope and Limitations

The scope of the research is limited to developing an earth observation data system for Murang’a County with focus on identifying patterns in urban development and monitoring the forest cover. The study attempts to address the key technical barriers that hinder the accessibility and availability of spatial data for decision making purposes. Mainly, open-source technologies would be utilized and open-access earth observation data will be utilized in order to keep the costs at a minimum while also delivering an effective and functional system.

Despite the importance and advantage of using the open-access data to reduce costs, this freely available data is constrained in terms of coverage and resolution, this in turn will affect the precision of the deforestation and urban development analyses. Additionally, geospatial data is large in volume and inherently complex thus the process of development of the necessary data pipelines and automating them is a significant technical challenge. Murang’a County’s characteristics may limit the direct application of some of the solutions intended to be developed for Murang’a to be used elsewhere. These challenges help in maintaining a fixed scope for the research.

Chapter 2: Literature Review

2.1 Introduction

Earth observation data has become a valuable resource in spatial monitoring and monitoring the environment. EO data is gathered from satellites and other remote sensing platforms and offers dependable information about phenomenon on the surface of the earth over different periods of time for purposes of analyses. According to Gomes et al. (2021), the accessibility, availability and consistency of EO data has made it useful in multiple applications on a global scale. These applications are such as evaluation of environmental issues such as land degradation, land misuse, deforestation and even disasters such as flood hazards. EO data helps us understand these phenomena even better and aids us in taking adequate measures to manage the outcome of the phenomena.

Nationally, EO data has grown in its applications in the public and private sector with entities using the data in environmental conservation and disaster mitigation initiatives. An example is predictive modeling using EO data which enables better monitoring of flooding and allows for interventions to be undertaken to reduce the damage (International Water Management Institute, 2024). Furthermore, EO data offers insights into environmental conditions and land-use patterns which supports sustainable urban planning by assisting policymakers make informed decisions which result in positive impact on the environment and populace. Through the use of EO data, stakeholders in the environment and spatial planning sectors are better equipped in predicting environmental issues and implementing countermeasures. In this way, the utilization of EO data is clearly useful in sustainable development in the country.

2.2 Empirical Literature Review

2.2.1 Volume and Variety of Data

EO data analysis regularly involves management of massive volumes of data from multiples sources such as sensors and satellites. There are a number of satellite platforms offering satellite data on an open-access license such as Sentinel and Landsat satellites. These are satellites that have been in orbit for years and decades continuously capturing images of the earth and transmitting them to ground stations. Years and years of data translates to a significant volume of image archives. According to Hu et al. (2018), the ever-increasing amount of EO data raises the demand for reliable data structures to effectively handle the size of these datasets.

Lenka et al. (2016) emphasizes the usefulness of tools such as Spatial Hadoop and GeoSpark in processing massive amounts of geospatial data due to their ability to deal efficiently with structured and unstructured data. As an example, the Swiss Data Cube utilized images from the Landsat series of satellites and has 3,386 images that cover Switzerland which totals to 867.5 GB of data covering the period from 1984 to 2017. This is a significant volume while also taking into consideration that there was missing data from some periods between 1991 and 1998. It is also key to consider that the size of Switzerland is 41,285 km^2 in comparison to Kenya which is 582,646 km^2. Therefore, for this research, it is considerable to consider a county such as Murang'a which is 2,559 km^2 in size.

2.2.2 Data Formats and Standards

Geospatial data is available in several formats and analysis of these different formats is important in exploiting the potential in the use of this data. These formats include common formats like Hierarchical Data Format (HDF) used for complex satellite data storage, Network Common Data Format (NetCDF) which is useful for climate and atmospheric data, GeoTIFF which works well for satellite imagery and aerial photos with geographic metadata and Gridded Binary (GRIB) which is used for meteorological forecasts. Each format has its own advantages such as HDF is excellent in handling multidimensional arrays, NetCDF excels at time series data, GeoTIFF is excellent for capturing raster data with spatial information and finally, GRIB zips together weather prediction data effectively.

In handling this multiplicity of data, Li et al. (2016) proposes the creation of geospatial web services to optimize the benefits of spatial data since such services and offer interoperability across heterogenous datasets. Additionally, the use of standards such as the Open Geospatial Consortium (OGC) are important in enabling the ready sharing and use of data across platforms. Table 2.1 showcases the different data collected by the Swiss Data Cube from the Landsat and Sentinel platforms. This demonstrates the usefulness of data cube architecture in integrating the NetCDF data format.

Table 1.1: ARD Collections Stored in SDC (Chatenoux et al ,2021)

Name	Platform	Product Type	Measurements	Description	CRS	Format
s2_l2a_10m_swiss	SENTINEL-2	dc_preproc	coastal_aerosol, blue, green, red, veg5, veg6,	Standard surface reflectance	EPSG:4326	NetCDF

			veg7, nir, narrow_nir, water_vapour, swir1, swir2, slc	related bands		
ls5_ledaps_swiss	LANDSAT-5	LEDAPS	blue, green, red, nir, swir1, swir2, pixel_qa, radsat_ qa, cloud_qa	Standard surface reflectance related bands	EPSG:4326	NetCDF
ls7_ledaps_swiss	LANDSAT-7	LEDAPS	As Landsat 5	Standard surface reflectance related bands	EPSG:4326	NetCDF
ls8_lasrc_swiss	LANDSAT-8	LaSRC	As Landsat 5	Standard surface reflectance related bands	EPSG:4326	NetCDF
s1_l3comp_swiss	SENTINEL_1_L3C	Gamma0	VV, VH	Radiometrically normalised (terrain-flattened) backscatter	EPSG:4326	NetCDF

The Swiss Data Cube also incorporates a number of steps in preparing the data before presenting it to the user as analysis ready data. The steps are shown in the processing workflow in Figure 2.1 giving a visual representation of the different levels that the data goes through. The user raises a request for data and the system searches the varied cloud service providers such as Google Earth Engine and Amazon Web Services hosting images from satellite image distributors. Upon discovery, the data is accessed and pre-processed and stored for future access and dissemination to users. This illustrates one of the methodologies utilized in accessing and processing the different formats of EO data that are available from different service providers.

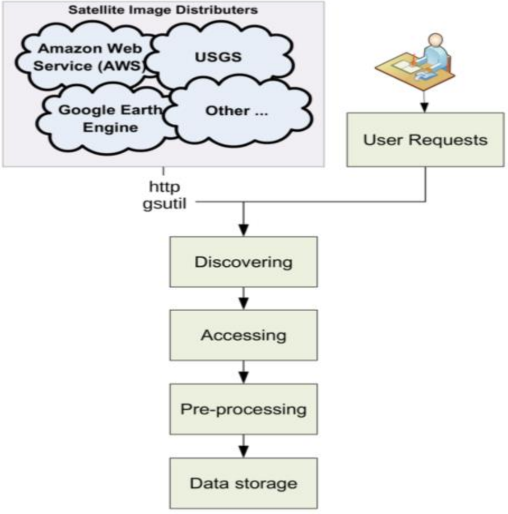


Figure 2.1: EO Data Processing Workflow (Guiliani et al, 2017)

2.2.3 County-Level Case Studies

The potential use of EO data across counties in Kenya for environmental management and disaster mitigation is high. Research by Omondi (2022) shows that EO data is useful in an integrated approach in developing Murang’a county’s capacity in pointing out high risk areas and develop early warning systems thus improving the livelihoods of the county residents. Similarly, Makueni County also showcases a distinct use of EO technology in its agricultural management initiatives which has improved sustainable practices. The county is reported to utilize satellite imagery in its mango cultivation. Another county that has demonstrated the capability of EO data in understanding complex environmental relationships is Elgeyo Marakwet County. Kilimo’s (2014) research used satellite imagery to show evidence of relationships between land cover changes and landslide occurrences in the county. The research showed that deforestation and drastic land use patterns affected the risk of disasters occurring. This analysis showcased the usefulness of EO data in influencing decision making and policy making for improving the quality of life of residents across counties.

Kenya Forestry Research Institute (2018) reports that satellite imagery analysis has been useful in precision monitoring of land use patterns and evaluating agroforestry initiatives in Kenya. Kenya News Agency (2023) captures a report acknowledging the potential of EO data to address food security challenges by supporting agriculture. The barriers identified in exploiting the potential of EO data is the limited technical capacity and inadequate inter-agency collaboration in the sharing of data and expertise. This analysis of counties and agencies such as Kenya Forestry Research Institute show that there isolated cases that show the usefulness of

EO data. However, there is no integrated platform that provides accessibility and availability of EO data.

2.3 Theoretical Literature Review: Data Cubes Concept

The Open Data Cube (ODC) concept is a significant innovation in the management of EO data as it moves from the traditional scene-base approach to a complex pixel grid model which allows for more indepth and detailed analysis. The Earth Observation Open Data Cube (EODC) represents a transformative advancement in Earth observation data management, transitioning from traditional scene-based approaches to a more sophisticated pixel grid model. This innovative framework substantially improves data accessibility and analysis capabilities through its comprehensive four-layer architecture: data acquisition, data cube infrastructure, application platform, and user interface. The system's implementation of analysis-ready data (ARD) significantly reduces the complexities traditionally associated with large-scale spatiotemporal data processing.

The EODC's technical infrastructure leverages powerful open-source technologies, including GDAL, Xarray, and Numpy, enabling robust data format compatibility and sophisticated pixel-level processing capabilities. According to Cao, Li, Yao, & Ma (2022), this architecture supports diverse operational environments, ranging from local systems to cloud platforms and high-performance computing facilities. The framework's flexibility allows researchers and practitioners to implement customized analytical algorithms and applications, effectively addressing the limitations inherent in conventional satellite imagery management systems.

Despite these significant advantages, the EODC's sophisticated nature presents certain implementation challenges. The system demands considerable technical expertise, requiring users to possess advanced skills in installation, configuration, and programming. This technical barrier to entry can impede widespread adoption, particularly among users lacking specialized computational knowledge. These implementation challenges highlight the need for continued development of more user-friendly interfaces and simplified deployment processes to broaden the EODC's accessibility across diverse user groups.

2.3.1 Frameworks and Models

A number of models and frameworks have been created to efficiently use EO data. The Swiss Data Cube, for instance, offers a methodology that arranges EO data into a logical structure for analysis, increasing user accessibility (Chatenoux et al., 2021). These frameworks make it easier to incorporate EO data into current systems, enabling more thorough studies and encouraging cooperation between the various parties engaged in spatial planning and environmental monitoring.

A robust open-source platform for organizing, processing, and evaluating massive amounts of Earth Observation (EO) data is the Open Data Cube (ODC) architecture. Along with strong metadata management and data provenance monitoring, it offers a Python-based API that makes it possible to efficiently classify, organize, and analyze large EO datasets. Applications ranging from local to continental-scale analyses can benefit from the ODC framework's scalable processing capabilities, flexible data access, and multi-sensor data integration. When working with large geospatial data, it provides capabilities like query optimization, on-demand processing, and lazy evaluation to improve speed. The framework is a flexible tool for scientists, academics, government organizations, and companies dealing with Earth observation data because of its architecture, which enables implementation on a variety of environments, including cloud platforms and high-performance computer clusters.

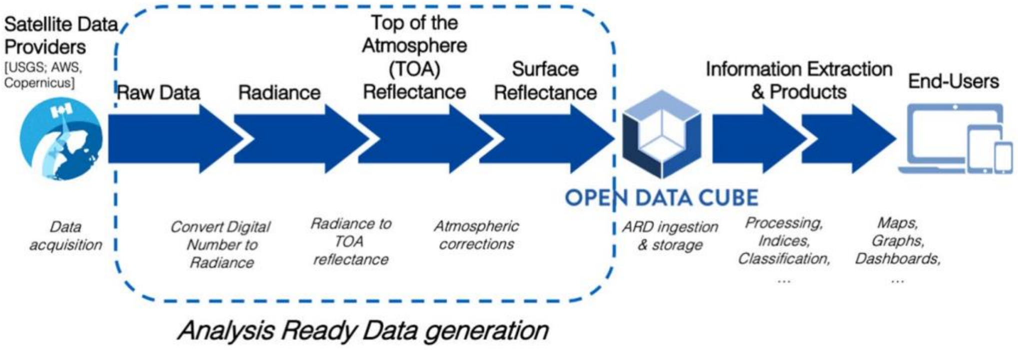


Figure 2.2: Common Workflow for Analysis Ready Data (Chatenoux et al, 2021)

An automated system for managing Earth Observation data, the Live Monitoring of Earth Surface (LiMES) framework was created to expedite data access, discovery, and pre-processing into Analysis Ready Data products. It creates a scalable and effective analysis system by combining high-performance computation, interoperable standards, and massive storage capacity. This is essential for creating complete data cube solutions like the Swiss Data Cube. The Live Monitoring of Earth Surface (LiMES) architecture has been used to generate Analysis Ready Data (ARD) products from a variety of satellite sources, such as Landsat 5, 7, 8, and Sentinel-1 and Sentinel-2 (see Figure 2.2). With its system of decomposable chains of interoperable services, this architecture makes it easier to find, access, and preprocess Earth observation data automatically. The LiMES framework improves the Swiss Data Cube's (SDC) capacity for thorough environmental analysis by effectively transforming unprocessed EO data into valuable environmental monitoring information products.

A comparison of several studies devoted to the creation and application of Earth observation (EO) data cubes is shown in Table 2.1. This table highlights the variety of data sources used by summarizing the main characteristics, applications, and technology utilized in each study. Chatenoux et al. (2021) employ Sentinel and Landsat data to build an analysis-ready data library for land use planning and environmental monitoring in Switzerland. Cao et al. (2022) concentrate on massive EO data management techniques that serve a variety of resource management and environmental monitoring applications. Gomes et al. (2021), on the other hand, highlight the processing and availability of EO data in Brazil, leveraging the Open Data Cube platform to promote interoperability. The table is a useful tool for comprehending the various methodologies and technology frameworks used in the analysis of EO data in several geographic situations.

Table 2.1 : Analysis of Data Cubes

Study	Data Sources	Key Features	Applications	Technologies Used
Chatenoux et al. (2021)	Landsat, Sentinel-1, Sentinel-2	Analysis-ready data archive; integrates various EO datasets for comprehensive analysis.	Environmental monitoring and land use planning in Switzerland.	Open Data Cube infrastructure, cloud-based storage.
Cao et al. (2022)	Various EO datasets	Designed for big EO data; emphasizes practices and lessons learned in managing extensive datasets.	Supports diverse applications in environmental monitoring and resource management.	Open Data Cube framework, cloud computing technologies.

Study	Data Sources	Key Features	Applications	Technologies Used
Gomes et al. (2021)	Landsat, Sentinel, CBERS	Facilitates access and processing of EO data using the Open Data Cube platform; focuses on interoperability.	Land use change detection, deforestation monitoring, and resource management in Brazil.	Open Data Cube platform, cloud-based processing tools.

2.3.2 Architectures and Designs

The functionality and usability of EO data systems are greatly influenced by their architectural designs. In the framework of EO data architectures, Lewis et al. (2019) explain the idea of Analysis Ready Data (ARD), which simplifies data processing and improves usability for a range of applications. Organizations can improve responsiveness in areas like urban planning and crisis management by ensuring that EO data is easily accessible for analysis through the establishment of a strong architecture.

The general architecture and software components of the Swiss Data Cube (SDC), as described by Chatenoux et al. (2021), are depicted in Figure 2.3. The different interrelated components that make it possible to organize and analyze Earth observation data effectively are summarized in this graphic. The user interface for accessing and displaying the data, the processing algorithms for producing Analysis Ready Data (ARD), and the automatic ingestion mechanism are important parts. The architecture is made to make it easier for various services to integrate and work together, which increases the usefulness of EO data for environmental monitoring and decision-making.

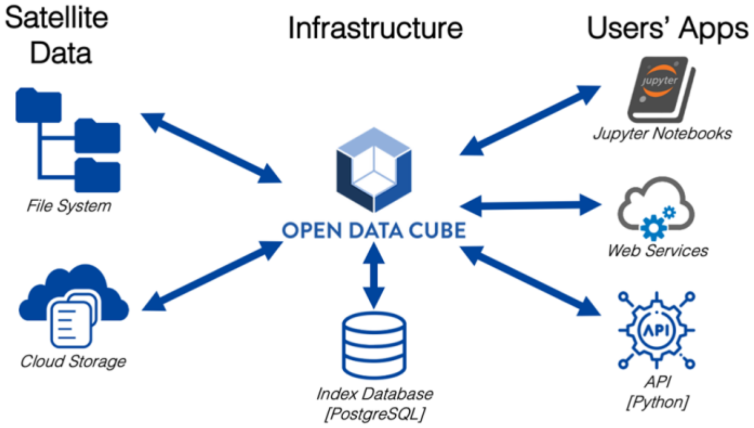


Figure 2.3: Swiss Data Cube General Architecture (Chatenoux et al (2021)

The architecture of the China Data Cube (CDC), including its structural elements and data flow procedures, is depicted in Figure 2.4. The integration of many Earth observation datasets and the techniques used for efficient data management and analysis inside the CDC framework are highlighted in this graphic..

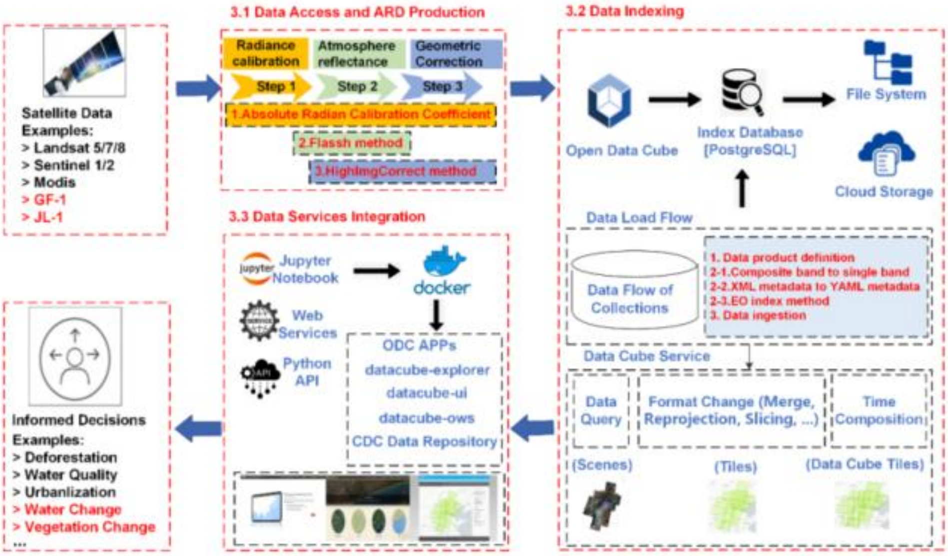


Figure 2.4: China Data Cube Architecture

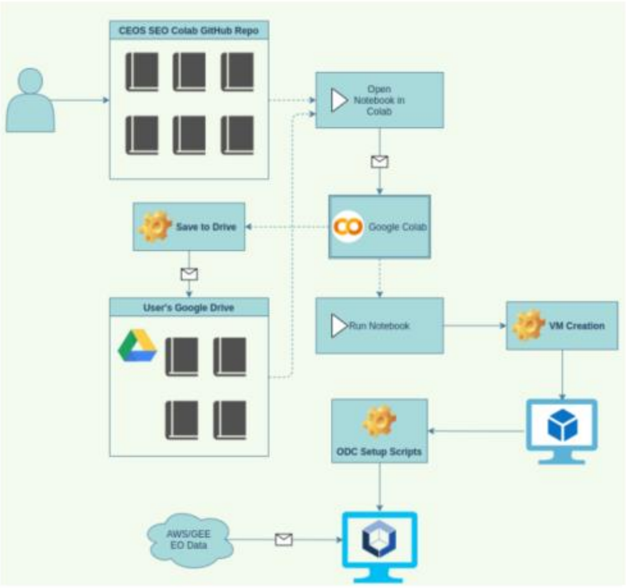


Figure 2.5: ODC CoLab Architecture (Killough et al, 2019)

The Open Data Cube (ODC) Colab architecture diagram, shown in Figure 2.5, describes the structure intended for cooperative analysis of Earth observation data. In order to facilitate effective access and analysis of geographic data, this diagram shows how different elements of the ODC ecosystem—such as data input, processing pipelines, and user interfaces—are integrated. According to Killough et al. (2019), the architecture facilitates research and decision-making processes in resource management and environmental monitoring by supporting a variety of applications.

2.4 Algorithms

The analytical capabilities of Earth Observation (EO) data cube systems depend fundamentally on their underlying algorithmic frameworks for data processing and interpretation. Advanced machine learning methodologies serve as the cornerstone for extracting meaningful insights from extensive EO datasets, enabling sophisticated pattern recognition and trend analysis across temporal and spatial dimensions.

Research by Cao et al. (2022) demonstrates the particular effectiveness of supervised classification algorithms in EO data analysis, with Support Vector Machines (SVM) and Random Forests emerging as primary tools for land cover classification and environmental change detection. These supervised learning approaches excel in their ability to process complex geospatial data while maintaining high accuracy in classification tasks. The implementation of unsupervised clustering algorithms, particularly K-means clustering, provides complementary analytical capabilities by identifying spatial patterns such as deforestation hotspots without requiring predefined classification parameters

2.4.1 Temporal Analysis and Insights

Time-series analysis algorithms serve as essential tools for extracting temporal patterns from Earth Observation (EO) data, enabling sophisticated tracking of environmental changes across different timescales. As documented by Cao et al. (2022), these analytical approaches excel in identifying both seasonal variations and long-term environmental trends, providing critical insights for strategic decision-making in urban development and disaster management sectors.

The integration of these advanced machine learning algorithms within the data cube framework enhances organizations' capabilities to process complex spatial data effectively, generating actionable insights for evidence-based decision-making. Figure 2.6 demonstrates this practical application through the visualization of Australia's annual land cover changes, illustrating how time-series analysis transforms raw EO data into meaningful environmental intelligence. This implementation showcases the practical utility of temporal analysis in generating comprehensive environmental assessments that support strategic planning and resource management decisions.

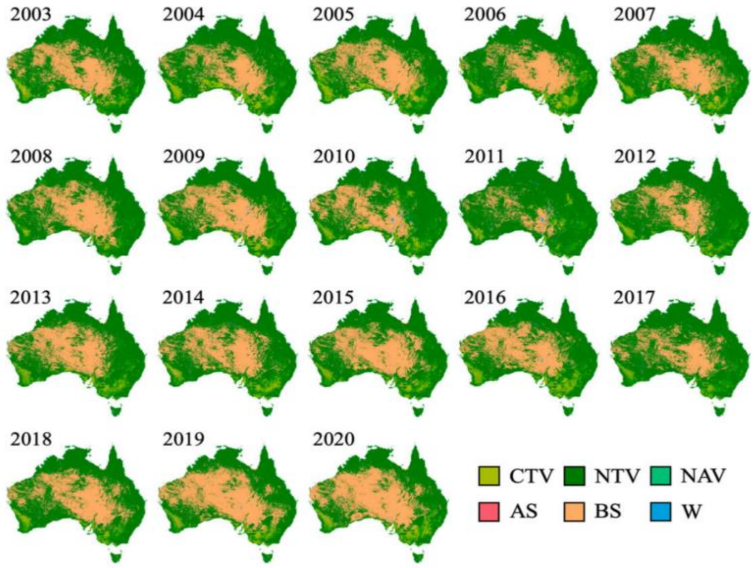


Figure 2.6: Annual Land Cover Maps for Australia (Owers et al, 2022)

2.4.2 Atmospheric Correction and Feature Extraction

Atmospheric correction and feature extraction algorithms play a fundamental role in Earth Observation data analysis. Two key algorithms, the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) and Land Surface Reflectance Code (LaSRC), are instrumental in correcting atmospheric disturbances in Landsat imagery. These algorithms specifically address challenges such as aerosol scattering and thin cloud presence, enabling more accurate data interpretation (Chatenoux et al., 2021). The ability to perform these corrections is particularly crucial for time-series analysis, as it ensures meaningful comparisons across temporal datasets. Furthermore, the Function of Mask (Fmask) algorithm has significantly advanced the detection capabilities for clouds, cloud shadows, and snow cover in both Landsat and Sentinel-2 imagery. This enhancement in detection accuracy has substantially improved the overall quality and reliability of Earth Observation products, making them more valuable for scientific and practical applications (Chatenoux et al., 2021).

2.5 System and Applications

The systems created to handle EO data are essential to its use in many different industries. Environmental monitoring, land-use planning, and disaster response can all benefit from the Open Data Cube platform's extensive framework for accessing and interpreting EO data (Gomes et al., 2021). These uses demonstrate how adaptable EO data systems are in tackling intricate problems pertaining to resource management, urban development, and environmental sustainability, ultimately fostering more resilient communities. The programs utilized in the Brazilian Data Cube System's deployment are displayed in Figure 2.7.

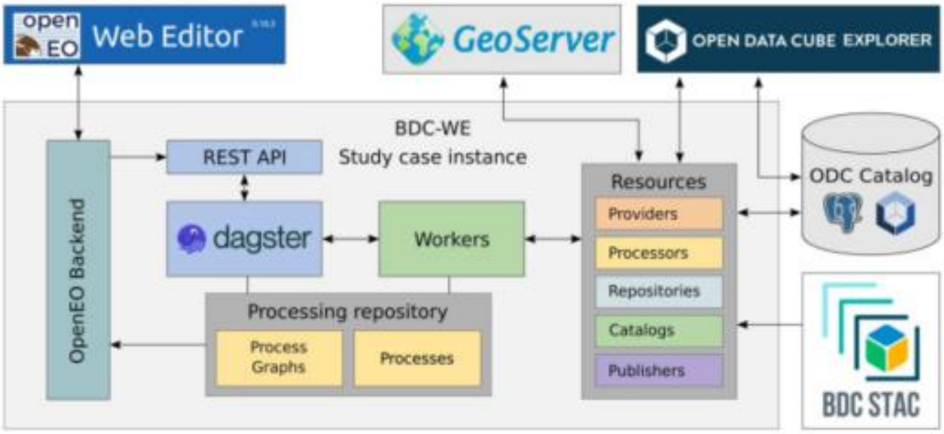


Figure 2.7: Brazilian Data Cube System (Gomes et al, 2021)

2.5.1 Data Cube Frameworks

In order to efficiently handle and evaluate the enormous volumes of satellite imagery and associated data, the creation of Earth observation data cubes requires a full suite of technologies and applications. The Data Cube Framework, like the Open Data Cube (ODC), is one of the fundamental elements. It arranges satellite imaging files for particular regions across predetermined time periods. Applications like environmental monitoring and land use analysis benefit greatly from the integrated gridded data analysis environment that tools like ODC offer, which enables users to effectively examine changes across a variety of landscapes. These frameworks simplify the data distribution process by offering analysis-ready data (ARD), which drastically cuts down on the time and effort needed for data preparation (Cheng et al., 2020; Dhu et al., 2019).

2.5.2 Data Integration Tools

Data Integration Tools are another essential component that makes it easier to transform various data formats into a consistent structure that can be analyzed inside a data cube. By linking different data sources and providing easy access to datasets that are ready for analysis, open-source solutions such as Cube.js can act as an abstraction layer that streamlines data integration (Cube.js, 2021). By guaranteeing interoperability across diverse datasets, these tools enable researchers to carry out in-depth analyses free from technological obstacles. In industries including agriculture, urban planning, and disaster management, the integration of several data sources improves the usefulness of Earth observation data and facilitates better decision-making (Li et al., 2016).

2.5.3 Cloud Computing Infrastructure and User Interfaces

Data cube performance and scalability are greatly aided by cloud computing infrastructure. Users may access and analyze data in real-time thanks to platforms like Amazon Web Services (AWS), which make it possible to store and process massive information efficiently. To improve functionality, open-source cloud technologies can also be incorporated. For dynamic applications where prompt insights are essential, such environmental monitoring and catastrophe response, this capacity is extremely advantageous (Shao et al., 2017). Because the cloud architecture offers a common platform for data access and analysis, it also makes it easier for different stakeholders to collaborate.

Furthermore, to guarantee that non-technical users may readily access and evaluate the data these systems give, user-friendly interfaces are required. Tools like Cube.js, which provide simple connection with front-end frameworks to produce unique visualizations, can be used to construct intuitive dashboards (Cube.js, 2021). These user interfaces make it easier to query and visualize intricate datasets, increasing the accessibility of Earth observation data for researchers, practitioners, and policymakers (Cao et al., 2022). Because of its simplicity, satellite data may be used more widely, which promotes innovation across a range of industries. Like Figure 2.8 from the Swiss Data Cube, the KDC's development should also concentrate on making user-friendly interfaces and reducing the technical know-how needed for its implementation.

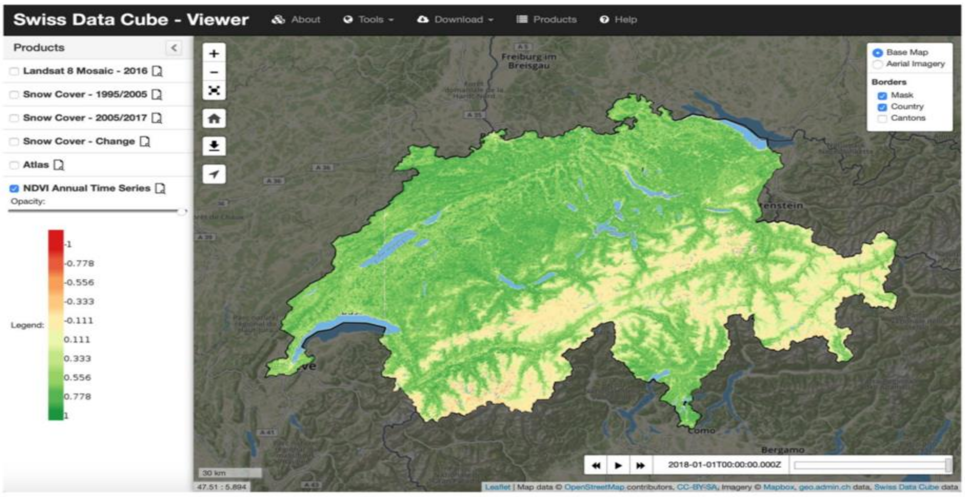


Figure 2.8: An Example of Swiss Data Cube (Chatenoux et al, 2021)

2.5.5 Machine Learning Algorithms and Real-Time Processing Capabilities

Additionally, analytical skills are improved by incorporating machine learning algorithms into data cube frameworks. Advanced analyses like predictive modeling and change detection, which are essential for applications in resource allocation and environmental management, can be implemented using open-source libraries like TensorFlow or Scikit-learn. Machine learning may greatly increase the accuracy and efficiency of geospatial studies by automating complicated activities (VoPham et al., 2018).

Lastly, in contemporary Earth observation applications, real-time processing capabilities are becoming more and more crucial. For efficient disaster response and urban planning, systems built to handle real-time data inputs provide for instant insights based on current conditions. Real-time data processing and streaming can be facilitated by open-source frameworks such as Apache Kafka. Rapid data processing facilitates prompt decision-making in urgent circumstances (Soliman & Terstriep, 2020).

2.6 Research Gap

There are still large research gaps, especially in Kenya, despite improvements in the accessibility and analysis of Earth Observation (EO) data. Natural resource management, disaster response, and efficient urban planning are all hampered by the inadequate incorporation of EO data into local decision-making procedures. It is essential to methodically gather and incorporate variables including high-resolution satellite images, land-use classifications, flood risk indicators, and socioeconomic data into an extensive data cube system in order to close these gaps. By ensuring that it satisfies the unique requirements of stakeholders across multiple sectors, this integration will increase the relevance of EO data.

Furthermore, EO data's utility is hampered by the absence of defined procedures for local processing and interpretation. To increase this data's usefulness and efficacy, uniform methods for handling it must be developed. EO data can be more effectively used to inform and support proactive management efforts by recognizing and resolving these issues. In the end, this will result in better decision-making procedures and better outcomes for Kenya's natural resource management, urban planning, and disaster response.

High-resolution satellite imagery, land-use classifications, flood risk indicators, socioeconomic data, and changes in the urban footprint over time are all examples of independent variables in this study. These factors have an impact on dependent variables including the success of urban design, the effectiveness of disaster response, the results of natural resource management, and indicators of socioeconomic progress. The study intends to evaluate how EO data might improve management strategies and decision-making processes in Kenya's many sectors by looking at these links.

2.7 Conceptual Framework

To optimize the usefulness of EO data, a structured approach integrating several components and approaches will be used in the proposed construction of a data cube system. Finding high-resolution EO data from satellite repositories will be the first step in the data gathering process, which will guarantee data quality and applicability for certain uses. The data will then be prepared for analysis by applying data preprocessing techniques like cloud masking and atmospheric correction. The EO data will then be arranged into multi-dimensional arrays by the data cube, which will enable effective retrieval and analysis in a data storage architecture. Like the Swiss Data Cube, the data will be aggregated from several sources, including Sentinel and Landsat satellite data.

Users will then be able to extract insights from the data through the integration of time-series analysis, clustering, and classification algorithms into the system. In order to enable stakeholders to engage with the data and perform intuitive analysis of the findings, the framework will also include user interfaces and visualization tools. For ongoing development, a feedback system will be put in place that will let users offer suggestions on the usefulness and relevancy of the data, which can inform further iterations of the data cube system. In addition to improving data accessibility, this conceptual framework will endeavor to empower local stakeholders by giving them the knowledge and resources they need to make wise decisions about the environment and space.

The creation of a Kenyan Data Cube will concentrate on resolving issues with data transmission, processing, and storage in order to effectively handle the expanding amount of Earth Observation (EO) data in Kenya. The system will make use of the openly available Open Data Cube (ODC) framework, which manages global analysis-ready data (ARD) and supports local infrastructure. The KDC will tackle particular issues in the Kenyan context, like handling data from multiple sources of satellite imagery in various formats to satisfy local project requirements. The KDC intends to create a localized cloud infrastructure that guarantees scalability and flexibility by learning from the China Data Cube experience. In order to maximize data storage and computing efficiency, future studies will investigate the integration of cloud platforms such as Microsoft Azure, Sentinel Hub, and Google Earth Engine, performing quantitative analysis.

The ODC framework has a lot of potential, but its manual configuration and need on programming knowledge may prevent it from being widely used in Kenya. As a result, attempts will be made to streamline installation procedures, offer thorough documentation, and integrate training programs. This will make it possible for Kenyan researchers, organizations, and decision-makers to use the KDC efficiently for a range of spatial data analysis and decision-making applications.

Earth observation data may be accessed and analyzed with ease thanks to the proposed Open Data Cube (ODC) technology. The ODC system will effectively process a request made by a user via a web interface.

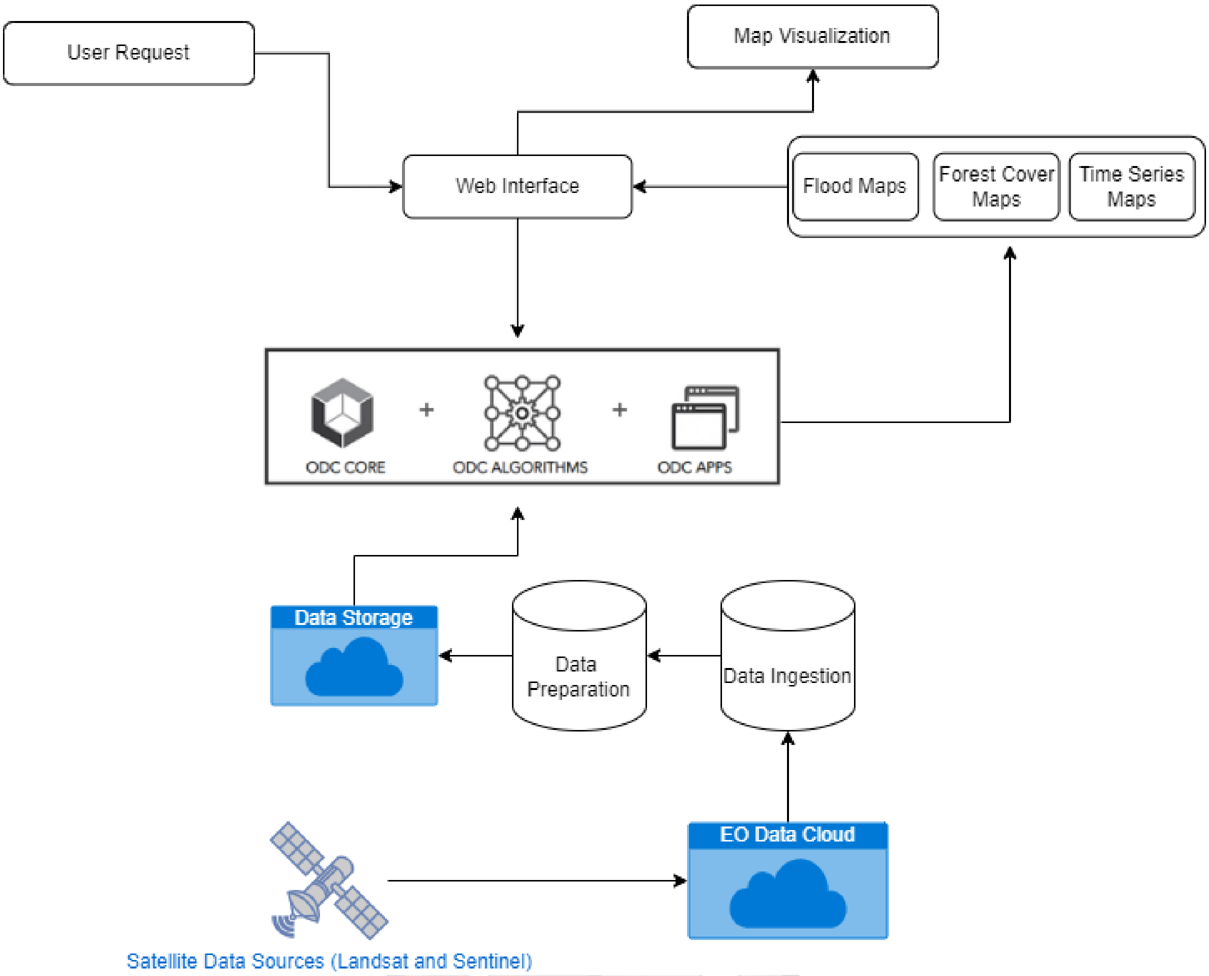


Figure 2.8: Kenya Data Cube Conceptual Framework

Images of the Earth will be taken by Earth observation satellites such as Sentinel and Landsat, analyzed, and stored in the cloud. The geographical Kenyan data will be ingested by the Kenya Data Cube (KDC) from this cloud storage, where it will undergo data correction and processing before being stored in its own cloud architecture. This processed data will be indexed by the ODC core for easy access. When a user request is received, the system will deconstruct it, retrieve the pertinent indexed data, and employ the necessary algorithms and applications from its database. The requested map will then be created by the ODC and returned to the user via the web interface. Users will not have to deal with the underlying data processing and storage issues thanks to this simplified procedure, which will make it simple to obtain and evaluate complex Earth observation data.

Chapter 3: Research Methodology

3.1 Introduction

Large amounts of Earth observation data can now be managed and analyzed with the help of data cubes. Several stakeholders must work together and prepare carefully in order to create a successful data cube. High-quality, analysis-ready data (ARD), which is usually obtained from satellite imaging, aerial photography, and ground-based sensors, forms the basis of a data cube. To guarantee uniformity in format, projection, and resolution, this data needs to be pre-processed. Involving a wide range of experts is essential to creating a solid data cube. These experts should include data scientists, remote sensing professionals, domain experts (such as those in forestry, agriculture, or urban planning), and end users from government organizations, academic institutions, and business. Their suggestions aid in defining the cube's structure, locating pertinent datasets, and guaranteeing that the finished result satisfies user requirements. Furthermore, interacting with local communities and legislators might yield insightful information about area goals and difficulties, which will aid in customizing the data cube for certain use cases and applications. Scalability, compatibility with current systems, and the possibility of future growth to accommodate additional data sources and analytical capabilities should all be taken into account during the development process.

3.2 Research Design

The best research design for creating an Earth Observation Data Cube (EODC) blends case study design and applied research design. While the Case Study component enables in-depth research of successful EODC implementations, the Applied Research component concentrates on developing novel technologies and finding solutions to real-world issues. This combination makes it possible to thoroughly examine user interface design, algorithm development, and data management in the context of practical applications. In the end, this strategy produces an EODC that can effectively handle pressing environmental issues and is both operationally and scientifically sound.

3.2.1 Applied Research Design

In essence, the creation of an Earth Observation Data Cube (EODC) is an applied research endeavor that tackles real-world issues in the administration and analysis of Earth observation data. For a number of important reasons, this strategy is especially appropriate for the goals of Kenyan Data Cube development. The problem-solving approach of applied research is well suited to the field's urgent requirement to handle and evaluate enormous amounts of satellite imagery data. Additionally, EODCs symbolize the applied research goal of creating state-of-the-art technologies by representing a novel technical way to handling Earth Observation data. The resulting EODC has a wide range of important practical applications, from monitoring land use and cover to detecting environmental change and responding to disasters. These practical uses highlight the importance of the applied research approach in developing instruments that have noticeable and instant advantages for Kenya's environmental management and decision-making procedures.

3.2.2 Case Study Design

Kenyan Data Cube development benefits greatly from incorporating case study design features in a number of important ways. Developers can improve their own implementation approach by studying successful EODC projects like the Swiss Data Cube, China Data Cube, and Brazilian Data Cube. This will give them important insights into potential problems and best practices. Additionally, case studies help focus the wide range of study into more focused, explorable instances, which is especially helpful while negotiating the intricate terrain of managing Earth observation data. More effective and focused research activities are made possible by this narrow emphasis. Additionally, case study design makes it easier to evaluate applications in the real world, allowing developers to determine whether particular theories and models apply to real-world events. In order to ensure that the finished product is both theoretically sound and practically feasible in the Kenyan setting, this practical component is essential when creating a working EODC system.

3.3 Methodological Approach

An Earth Observation Data Cube (EODC) system's development framework consists of a wide range of operational and technological elements that are directed by tried-and-true procedures and successful deployments. The first step in this method is methodical data preparation, with an emphasis on combining Analysis Ready Data (ARD) from several Earth observation platforms into a single framework. The design ensures scalability and cost-effectiveness in data processing and storage operations by utilizing cloud computing infrastructure to effectively manage large Earth observation datasets.

An efficient data cube structure for handling multidimensional Earth observation data is incorporated into the core system design, which is bolstered by sophisticated indexing and processing tools. Through integrated geospatial operators, this framework supports both batch and real-time processing needs. Machine learning techniques are also incorporated for improved forecasting and analytical capabilities. In order to guarantee data quality and processing correctness, the implementation places a strong emphasis on system interoperability through defined protocols, user-centric interface design, and thorough validation methods.

The resulting EODC architecture allows for the smooth integration of new data sources and analysis techniques by upholding adaptability as a basic design feature. This proactive strategy guarantees the system's ongoing applicability and efficiency in assisting with Earth observation applications in a variety of fields.

3.4 Population / Sampling

Monitoring forest cover and analyzing urban growth are two main application areas that are the focus of the Earth Observation Data Cube's data collecting approach. The technology gathers temporal information and high-resolution satellite pictures that are ideal for monitoring trends of urbanization and forest dynamics. This focused strategy maximizes system efficiency while guaranteeing that the data infrastructure efficiently supports decision-making for both urban planning and forest management authorities.

3.4.1 Population

The target audience for Kenya's Earth Observation Data Cube (EODC) system development includes important governmental and private sector players with expertise in environmental management and geospatial technology. This comprises members from each of the 47 county administrations as well as experts from more than 280 state corporations. This population's size shows the extensive institutional architecture needed for EODC to be implemented and run under Kenya's governance framework. Representation from both national and regional levels of environmental and technological decision-making is ensured by this deliberate population selection.

3.4.2 Sampling

Targeting important institutional stakeholders with specialized knowledge of environmental management and geospatial technology, the sampling methodology for Kenya's Earth Observation Data Cube (EODC) development uses a strategic convenience sampling strategy. Representatives from the Department of Resource Surveys and Remote Sensing (DRSRS), Kenya Forest Service (KFS), and Kenya Space Agency (KSA) are included in this framework for selected sampling. Three county governments—Murang'a, Vihiga, and Elgeyo Marakwet—are represented in the study to guarantee thorough regional representation.

LocateIT, a specialist geospatial technology company, participates to integrate private sector experience. The inclusion of crucial technical knowledge and a range of institutional viewpoints required for the effective development and application of EODC in the Kenyan context is guaranteed by this methodically planned sampling strategy.

3.4.3 Data Collection/Instrumentation

Primary and secondary research approaches are incorporated into the data gathering plan for creating Kenya's Earth Observation Data Cube (EODC). Structured questionnaires serve as the primary data collection method, enabling quantitative assessment of stakeholder requirements, preferences, and implementation challenges. The questionnaires are distributed to staff members from participating organizations, including government agencies, county administrations, and private sector entities, to gather comprehensive insights into operational needs and potential implementation barriers.

The thorough examination of existing project documents from involved organizations strengthens the study foundation. The historical background of current geospatial technology applications and environmental management techniques in Kenya is crucially provided by this secondary data review. In order to optimize the EODC system design and implementation approach and make sure the solution fills in recognized operational gaps while building upon proven effective practices, it is helpful to analyze prior initiatives and their results.

The development of an EODC system that successfully meets stakeholder objectives while preserving operational sustainability within the Kenyan environment is supported by this integrated approach, which guarantees a comprehensive grasp of both technical requirements and institutional issues through systematic questionnaire-based assessment combined with thorough document analysis.

3.5 System Development

An Earth Observation Data Cube (EODC) system will be built using a systematic process, leveraging open-source technology throughout. The system design will be based on the Open Data Cube (ODC) framework, which offers a solid foundation for handling enormous volumes of satellite images and geographic data (Giuliani et al., 2020). This approach, which prioritizes collaborative development and open-source principles, aligns with successful implementations like the Brazil Data Cube (BDC) project (Ferreira et al., 2020).

3.5.1 Data Preparation and Management

Existing frameworks for handling Earth observation data are expanded upon by the research approach for data management and preparation (Giuliani et al., 2020). Using automated retrieval procedures that follow best practices in Earth observation data management, the procedure starts with the capture of Sentinel-1 and Sentinel-2 satellite imagery via the Copernicus Open Access Hub (Strobl et al., 2023). This strategy guarantees constant data accessibility and quality, which is especially important for developing nations like Kenya. Implementing a strong pretreatment pipeline in accordance with techniques confirmed by earlier successful Data Cube system deployments is necessary for data preparation (Dhu et al., 2019). The Sen2Cor processor's atmospheric correction for Sentinel-2 data expands on accepted practices to guarantee radiometric accuracy (Ferreira et al., 2020). In order to preserve spatial precision in Earth observation data, the geometric correction and spatial registration procedures are created using tried-and-true techniques (Dwyer et al., 2018).

In order to follow best practices from currently operating systems, the management framework integrates a hierarchical data storage structure (Killough, 2019). This strategy, which follows architectural patterns that have proven effective in comparable geographic contexts, makes use of PostgreSQL with PostGIS extension for metadata management and MinIO object storage. These elements work together to provide effective data handling while preserving the scalability and dependability of the system.

3.5.2 Processing and Analysis Capabilities

The processing and analysis capabilities are developed using the Open Data Cube framework, incorporating recent advances in distributed computing for Earth observation data analysis (Cao et al., 2022). The system's architecture builds upon successful implementations of cloud-based processing platforms for satellite imagery (Gorelick et al., 2017), while adapting these approaches to the specific requirements of the Kenyan context. Core processing capabilities include multi-temporal analysis and spectral index calculations, implementing methodologies validated through previous research in environmental monitoring (Chatenoux et al., 2021). The system's processing chains are designed following established practices in Earth observation data analysis, with specific adaptations for regional requirements (Asmaryan et al., 2020).

The system implements comprehensive multi-temporal analysis capabilities for tracking land cover changes, building on methodologies validated by Giuliani et al. (2020). This includes advanced spectral index calculations such as NDVI, NDBI, and NDWI, following standardized procedures outlined by Ferreira et al. (2020). The time-series analysis component implements approaches validated by Cao et al. (2022), while statistical analysis tools are developed based on frameworks described by Strobl et al. (2023). These capabilities are integrated into a cohesive processing framework that enables efficient analysis of temporal and spatial patterns in environmental data.

3.5.3 User Interface and System Optimization

The capabilities of the data cube will be easily accessible to users using a web-based interface developed using open-source frameworks like Flask or Django (Giuliani et al., 2019). System performance will be maximized by the use of cloud computing technologies, parallel processing techniques, and containerization with Docker for scalable deployment (Rizvi et al., 2019). Throughout the development process, comprehensive documentation will be maintained current and knowledge will be disseminated to the open-source community. Real-world case studies will be used for testing and validation in order to guarantee the system's dependability and effectiveness for its intended usage in Kenya's diversified geography (Dhu et al., 2019).

3.5.4 Resource Planning and Cost-Effective Implementation Strategy

3.5.4.1 Computational Resource Options Analysis

When implementing an Earth Observation Data Cube system, computational resources must be carefully considered while budgetary and performance needs are balanced. Both local and cloud-based choices are shown in the study that follows, along with their corresponding costs and specifications.

Table 3.1: Local Workstation Specifications and Costs

Component	**Minimum Specification**	**Recommended Specification**	**Estimated Cost (KShs)**
CPU	8 cores, 3.6GHz	16 cores, 4.0GHz	52,000 - 104,000
RAM	32GB DDR4	64GB DDR4	26,000 - 52,000
Storage	4TB HDD + 512GB SSD	8TB HDD + 1TB NVMe SSD	39,000 - 78,000
GPU	8GB VRAM	12GB VRAM	65,000 - 104,000
Motherboard	Standard ATX	High-end ATX	26,000 - 39,000
Total Cost			**221,000 - 396,500**

Table 3.2 Cloud Computing Options and Monthly Costs

Service Type	Basic Tier	Standard Tier	Performance Tier
Compute (vCPUs)	4 cores	8 cores	16 cores
Memory	16GB	32GB	64GB
Storage	1TB	2TB	4TB
Monthly Cost (KShs)	**26,000 - 39,000**	**52,000 - 78,000**	**104,000 - 156,000**

3.5.4.2 Cost-Effective Open Source Implementation Strategy

This study suggests a hybrid strategy that makes use of open-source solutions and minimum cloud resources, based on the evaluation of available options and budgetary restrictions. The suggested approach makes strategic use of cloud infrastructure in conjunction with a wide range of open-source components.

Table 3.3: Recommended Open Source Software Stack

Component	Open Source Solution	Purpose	Cost
Operating System	Ubuntu Server 22.04 LTS	Base system	Free
Data Cube Framework	Open Data Cube	Core functionality	Free
Database	PostgreSQL + PostGIS	Spatial data management	Free
Processing Framework	GDAL/OGR	Geospatial processing	Free
Visualization	GeoServer	Data visualization	Free
Development Environment	JupyterLab	Analysis and development	Free
Container Platform	Docker + Kubernetes	Deployment and scaling	Free

Table 3.4 Minimal Cloud Resource Requirements

Resource Type	**Specification**	**Monthly Cost (KShs)**
Compute Instance	8 vCPUs, 32GB RAM	20,800
Object Storage	2TB	5,200
Data Transfer	500GB/month	5,850
Database Instance	4 vCPUs, 16GB RAM	10,400
Total Monthly Cost		**42,250**

3.5.4.3 Recommended Implementation Approach

The implementation plan uses a staged method that blends modest cloud resources with local development. The first phase focuses on setting up a local development environment using an entry-level workstation configuration. The entire stack of open-source software will be housed in this environment, allowing for the processing of smaller datasets, prototype analytical techniques, and the local development and testing of processing workflows.

In the second stage, rudimentary cloud infrastructure is deployed for data processing and storage, introducing basic cloud integration. This stage permits the processing of bigger datasets and time series analysis while preserving the local development environment for testing and development. Scaled implementation, the last stage, allows for full production workflows and extensive analysis capabilities by modifying cloud resources in accordance with verified needs.

3.5.4.4 Cost Optimization Strategies

Comprehensive optimization techniques are used in the implementation to save expenses without sacrificing system performance. Implementing tiered storage with automated data lifecycle management, compressing data that is rarely accessed, and storing just processed data in the cloud while archiving raw data on local storage systems are all examples of storage optimization.

Batch processing for big datasets and planning intense processing for off-peak times are the main goals of processing optimization. The system keeps a cache of frequently requested processing results and uses spot instances for non-time-critical processing. For effective resource consumption, resource management techniques include container-based deployment and the automatic shutdown of idle resources.

3.6 Data Analysis Methods

3.6.1 Quantitative Analysis

According to Dwyer et al. (2018), the quantitative analysis technique adheres to accepted frameworks for assessing Earth observation systems. The strategy uses metrics that have been verified by earlier studies and covers both system performance and data analysis capabilities (Chatenoux et al., 2021). According to Ferreira et al. (2020), performance metrics are created using methodologies that include detailed measurements of system response times for data retrieval operations, processing throughput evaluation for a range of analysis tasks, in-depth resource utilization analysis under various operating conditions, and a thorough assessment of storage efficiency across the system's data management components.

3.6.2 Analysis of Qualitative Data

By using techniques that have been proven effective by other studies, the qualitative analysis framework expands on well-established methods for assessing geospatial systems (Asmaryan et al., 2020). A thorough user experience evaluation process covering several assessment variables will be used in the analysis. The study will collect in-depth information about the platform's efficacy and usability through detailed questionnaires completed by system users. A wide range of stakeholders, including government representatives, urban planners, and environmental scientists, will participate in the questionnaire-based assessment to provide a comprehensive understanding of the system's applicability in various use scenarios.

Think-aloud exercises, task completion analyses, and systematic usability testing sessions with target user groups are all part of the evaluation methodology. This method, which has been verified by Giuliani et al. (2020), offers insights into users' mental models and decision-making processes while making it possible to identify both overt and covert usability issues. Throughout the system's development phase, the research will put in place a framework for collecting feedback continuously, enabling incremental changes based on user experiences and recommendations.

A systematic evaluation approach that looks at several facets of system functioning and performance will be used to carry out the system capacity assessment. This involves a thorough examination of the system's flexibility in responding to various use scenarios, using the methodology developed by Strobl et al. (2023). The evaluation will look at things like data compatibility, processing chain flexibility, and output format adaptation in order to determine how well the system integrates with current workflows in Kenyan institutions. As recommended by Dhu et al. (2019), the study will also carry out comprehensive reliability studies, recording system behavior under varied operational situations and user loads.

3.7 Expected Outcomes

The study intends to produce a number of noteworthy results that will advance our theoretical knowledge and real-world implementation of Earth observation systems in Africa. The main result will be a fully operational Earth Observation Data Cube system tailored for the Kenyan environment, building upon the framework developed by Cao et al. (2022). With a focus on tackling regional issues including inconsistent data availability and constrained computational resources, this system will show that it can manage several satellite data sources while supporting a range of environmental monitoring applications.

The study will yield verified approaches for processing and evaluating satellite data in the Kenyan environment, adhering to the methodological approach described by Ferreira et al. (2020). These approaches will explicitly tackle issues including enduring cloud cover, significant seasonal fluctuations, and computing resource limitations. As documented by Killough (2019), the development of these approaches will take into account the lessons acquired from successful implementations in comparable contexts.

The creation of a thorough framework for combining local datasets with data from satellites will be a noteworthy result. Building on the ideas presented by Giuliani et al. (2020), this framework will guarantee compatibility with current data infrastructure while improving the system's usefulness for local decision-making processes. In accordance with Chatenoux et al. (2021) evaluation frameworks, the study will also generate quantitative metrics illustrating the system's efficacy, performance, and dependability in assisting environmental monitoring and decision-making processes.

3.8 Dissemination of Results

In order to maximize the impact and accessibility of research findings across various stakeholder groups, the dissemination strategy uses a multifaceted approach. The research findings will be published in scholarly journals that concentrate on environmental monitoring and Earth observation, when effective models have been used in comparable situations (Killough, 2019). In order to lay the groundwork for future implementations in comparable contexts, these publications will describe the methodological advancements and technical implementations made possible by the research.

The study will follow Gorelick et al. (2017)'s best practices and generate thorough system documentation and technical implementation recommendations. In order to ensure that the system can be successfully maintained and reproduced, this documentation will include comprehensive deployment guidelines, maintenance processes, and troubleshooting techniques. For the benefit of both technical and non-technical stakeholders, the documentation will be made available via a variety of channels, such as institutional archives and online repositories.

A key element of the dissemination strategy will be knowledge transfer activities, which will include techniques that Asmaryan et al. (2020) have verified. Creating online tutorials and webinars, setting up cooperative meetings with stakeholder organizations, and holding training seminars for prospective users in Kenyan institutions are some of these activities. In accordance with the frameworks proposed by Strobl et al. (2023), the study will also create a knowledge-sharing platform that enables continuous contact and assistance for system users.

3.9 Utilization of Results

The usage plan aims to guarantee the research findings' long-term durability as well as their immediate practical use. The system will provide vital information for natural resource management, agricultural monitoring, and urban planning in the near future, supporting environmental monitoring and evaluation initiatives in Kenya. This strategy incorporates regional concerns emphasized by Asmaryan et al. (2020) and builds upon successful implementation models recorded by Cao et al. (2022).

The research's long-term effects will promote evidence-based environmental policymaking and regional climate change adaption strategies by strengthening Kenya's Earth observation capacity. As a model for comparable implementations in other nations, the research will lay the groundwork for increased Earth observation applications throughout Africa, adhering to the methodology developed by Dhu et al. (2019). Establishing long-term alliances with important stakeholder groups is part of the usage strategy, which guarantees ongoing system development and adaption to new requirements.

The findings of the study will be used as a foundation for expanding the system to incorporate more data sources and sector-specific specialized applications. Building on the ideas presented by Giuliani et al. (2020), this framework will allow for ongoing system development and adaption to satisfy changing user requirements. Additionally, the implementation will provide a platform for ongoing Earth observation research and development, encouraging creativity and knowledge generation in the field of environmental monitoring and evaluation.

In order to guarantee efficient use, the study will create explicit guidelines for system upkeep and development, including feedback systems that permit ongoing enhancement in response to user experiences and new needs. This strategy, which has been confirmed by Chatenoux et al. (2021), will contribute to the system's long-term viability and applicability while offering a strong basis for upcoming advancements in Earth observation applications throughout Africa.

Chapter 4: System Analysis, Design and Architecture

4.1 Introduction

The development of the Earth Observation Data Cube System involved the integration of satellite imagery processing techniques with spatial-temporal indexing to establish a functional analysis platform for environmental monitoring. The system was designed to handle data from multiple sources—primarily Sentinel and Landsat—while supporting the generation of vegetation and urbanization indices. The architecture emphasized modularity, enabling efficient ingestion, preprocessing, and retrieval of geospatial data. Interaction between system components was defined through structured workflows and illustrated using use case, sequence, and data flow diagrams. The design phase also incorporated both functional and non-functional requirements, detailing the system's core capabilities, performance expectations, and usability considerations within the context of Murang'a County.

4.2 System Analysis

System analysis involves identifying the goals, constraints, and technical specifications necessary for building a functional solution. In the context of this study, the objective was to develop a spatial-temporal data cube system capable of processing and analyzing satellite imagery to detect patterns of deforestation and urban development within Murang'a County. This section outlines the essential system requirements—both functional and non-functional—that guided the architecture and implementation of the solution, ensuring it meets the analytical needs of environmental monitoring and spatial planning stakeholders.

4.2.1 Requirement Gathering

To collect data for this study, a structured questionnaire was administered to selected stakeholders involved in Earth Observation (EO) data use across government institutions, academia, and the private sector. The aim was to assess current usage patterns, satisfaction levels, challenges, and institutional readiness for adopting an EO Data Cube system. Respondents included professionals from agencies such as the Kenya Space Agency, Directorate of Resource Surveys and Remote Sensing (DRSRS), county governments, universities, and regional organizations like UN-Habitat and Mohammed VI Polytechnic University.

The questionnaire captured both qualitative and quantitative data, focusing on EO data sources, frequency of use, spatial and temporal resolution needs, software tools, technical expertise, and perceived barriers. Respondents highlighted key applications of EO data in environmental monitoring, agriculture, urban planning, and disaster management. Most organizations rely heavily on open-access satellite data (e.g., Landsat, Sentinel, MODIS) and prioritize high spatial resolution (1–10 m) and frequent temporal coverage (weekly to monthly). Tools such as QGIS, Google Earth Engine, ERDAS IMAGINE, and Python are commonly used, with most institutions confirming sufficient internal expertise for EO data processing.

Data analysis involved cleaning and transforming responses into a structured format for synthesis. Emerging themes were coded and categorized to assess patterns in EO data usage, system challenges—such as limited access to high-resolution imagery, processing capabilities, and data quality—and training needs. These insights directly informed the functional requirements of the proposed EO Data Cube, ensuring that the system design aligns with stakeholder needs for timely, high-quality, and accessible EO-driven analytics. A total of 96 questionnaires were distributed, with 17 responses received.

Table 4.1 - EO Data User Survey Summary Table

Organization Type	Uses EO Data	Main Applications	Data Sources	Resolution Preference	Main Challenges	Interest in EO Data Cube
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	High (1 to 10m)	Cost	Interested
Gov/Agency	Monthly	Env, Urban, Agri, Disaster	Sentinel, Landsat	Medium (10 to 30m)	Cost, Processing	Very Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	High (1 to 10m)	Cost, Data Access	Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	Medium (10 to 30m)	Cost, Data Access	Very Interested
Private	Daily	Agriculture, Defense	Sentinel, Aerial, LiDAR	Very High (< 1m)	Prefer not to say	Very Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	High (1 to 10m)	Technical Capacity	Very Interested

Academic	Daily	Env, Urban, Agri, Heritage	Sentinel, Aerial	Medium (10 to 30m)	Access to High Res	Very Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	Medium (10 to 30m)	Data Quality, Cost	Very Interested
Gov/Agency	Daily	Forest Mgmt	Landsat, MODIS, Drone	Medium (10 to 30m)	Cost	Very Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	High (1 to 10m)	Processing, Infra	Very Interested
Gov/Agency	Weekly	Env, Urban, Agri	Sentinel, Landsat	High (1 to 10m)	Cost, Technical Capacity	Very Interested
Academic	Weekly	Env, Urban, Agri	Sentinel, Landsat	High (1â€“10m)	Cost, Technical Capacity	Very Interested
Gov/Agency	Daily	Env, Urban, Agri	Sentinel, Landsat	Medium (10â€“30m)	Cost	Very Interested
Academic	Monthly	Env, Urban, Agri	Sentinel, Landsat	High (1â€“10m)	Cost, Processing	Very Interested
Gov/Agency	Weekly	Env, Agri, Water	Sentinel, Landsat	Medium (10â€“30m)	Resolution	Interested
UN/Int'l	Daily	Urban planning	Sentinel, Landsat	Medium (10â€“30m)	Data Accessibility	Neutral

4.2.2 Functional Requirements

The design of the Earth Observation Data Cube System was informed by stakeholder needs, system objectives, and technical constraints. Functional requirements define the essential operations the system must perform to support satellite data ingestion, preprocessing, index computation, and user interaction. These include acquiring imagery, generating analysis-ready data, computing indices such as NDVI and NDBI, organizing outputs into a spatiotemporal cube, enabling time-series queries, and presenting results through visualizations and reports. The system must also incorporate user management to ensure secure, role-based access to its functionalities.

Table 4.2 Functional Requirements

Requirement	Description
Data Ingestion	System must download or access satellite imagery from open-access repositories.
Preprocessing Pipeline	Perform atmospheric correction, cloud masking, clipping, and normalization.
Index Computation	Compute NDVI, NDBI, and optionally NDWI for environmental monitoring.
Data Cube Construction	Organize imagery into a multidimensional structure indexed by space and time.
Time-Series Querying	Support spatial and temporal filtering for trend analysis.
Visualization and Reporting	Display map overlays, time-series charts, and export results as GeoTIFF/CSV.
User Management	Enable authentication and role-based access to different system functionalities.

4.2.3 Non-functional Requirements

The non-functional requirements define the quality attributes essential for the effective, secure, and sustainable operation of the system. These include performance, scalability, reliability, and ease of use—ensuring responsive data processing, support for larger datasets, and accessibility for non-specialist users. Maintainability and interoperability with standard

GIS tools are also key, alongside robust access controls to protect data integrity and user activity.

Table 4.3 Non-functional Requirements

Requirement	Description
Performance	System should return query results within acceptable time under normal load.
Scalability	Design must support extension to larger geographic areas or additional datasets.
Usability	User interface must be intuitive and usable by non-technical stakeholders.
Reliability	Ensure data integrity and fault tolerance across ingestion, storage, and analysis.
Maintainability	System should support modular updates and bug fixes without disrupting services.
Interoperability	Enable integration with existing GIS platforms (e.g., QGIS, ArcGIS, GEE).
Security	Protect user data and restrict unauthorized access to processing capabilities.

4.3 System Architecture

The architecture of the Earth Observation Data Cube System adopts a modular, three-tier design that separates data acquisition, processing and storage, and user interaction. The data acquisition layer is responsible for retrieving satellite imagery from open-access sources such as Sentinel and Landsat, filtering scenes by geographic extent and cloud cover. The processing and storage layer performs atmospheric correction, cloud masking, and computation of geospatial indices like NDVI and NDBI. Processed outputs are then organized into a multidimensional data cube indexed by space, time, and spectral bands, with metadata stored in a spatial database. The application layer provides analytical and visualization tools, allowing users to perform time-series queries, generate thematic maps, and export results through a web-based or desktop interface. This layered approach ensures scalability, maintainability, and ease of integration with existing geospatial platforms.

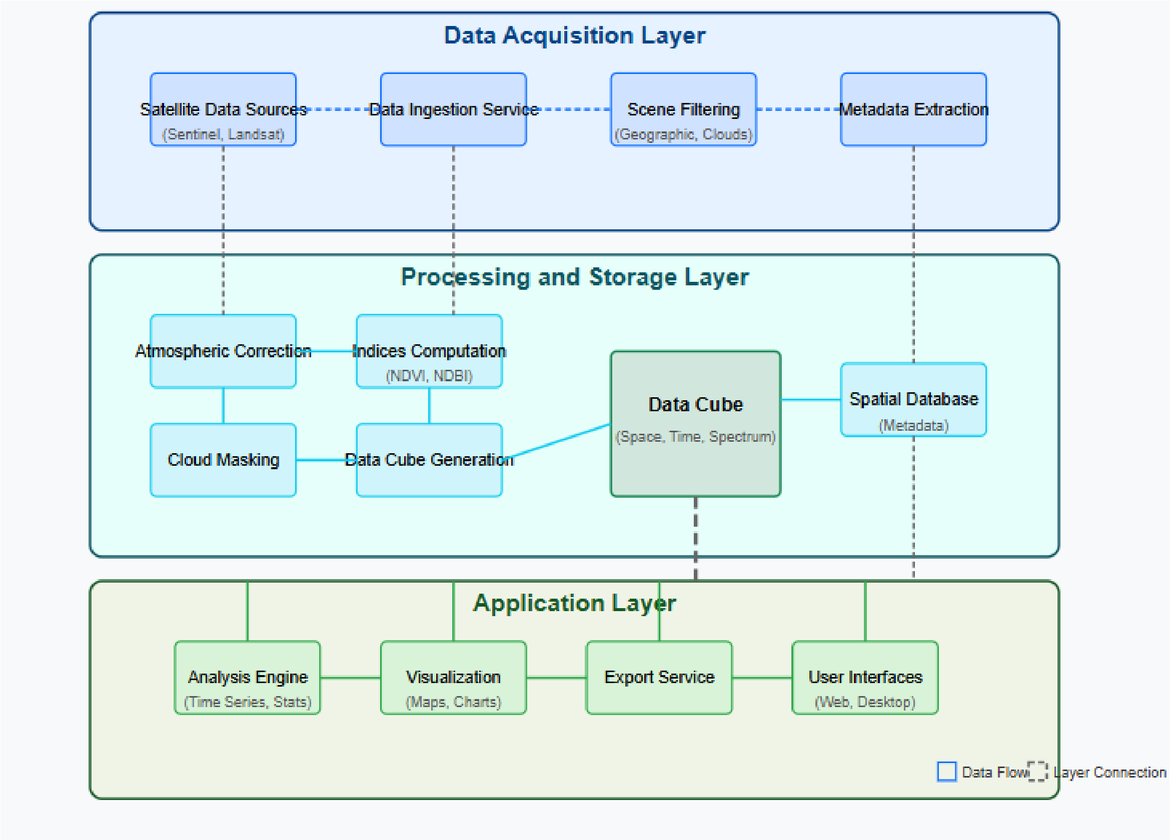


Figure 4.1 Earth Observation Data Cube System Architecture

4.4 System Design

In the System Design phase, the architecture and component interactions of the Earth Observation Data Cube System were formally defined using Unified Modeling Language (UML) diagrams. These models were informed by the findings from the system analysis and requirements gathering stages. A structured system design methodology was adopted to ensure a clear separation of functional components and data flow processes. The design emphasized modularity, allowing independent development and testing of ingestion, processing, storage, and visualization modules. Key artifacts included use case diagrams, sequence diagrams, context diagrams, and data flow diagrams—each outlining how users and system components interact to support the system's analytical and operational goals.

4.4.1 Use Case Diagram

The use case diagrams illustrate the functional interactions between external actors and the Earth Observation Data Cube System. They represent typical scenarios where users—such as GIS analysts, planners, and urban development officers—engage with the system to request NDVI or NDBI maps, analyze forest cover changes, or assess urban expansion trends. Each use case highlights a specific system function initiated by the actor, helping to clarify system behavior from the user's perspective. These diagrams provide a visual understanding of user-system relationships and inform the basis for designing user-centered workflows.

4.4.1.1 Forest Monitoring Use Case

The following use case describes a typical interaction between users and the system in the context of forest monitoring. It outlines how GIS analysts and planners utilize the system to generate NDVI-based forest cover change maps, including the conditions required, the expected outcomes, and the sequence of actions performed.

Table 4.4 Forest Monitoring Use Case

Use Case Name	Generate Forest Cover Change Map
Use Case Description	A GIS Analyst or Planner interacts with the system to request NDVI-based forest cover maps over a specific time period. The system processes the request and returns a visual output showing vegetation trends and areas of decline or improvement.
Actors	GIS Analyst, Planner
Pre-Condition	The system must have access to preprocessed satellite data with NDVI computed and indexed in the data cube.
Post-Condition	The user receives a forest cover change map along with summary statistics for the specified time range.
Main Success Scenario (Flow of Events)	The user logs into the system and selects the "Forest Monitoring" module. They define the spatial extent (e.g., county boundary) and temporal range (e.g., 2020–2023). The system queries the data cube, retrieves relevant NDVI data, performs temporal differencing, and generates a map highlighting forest gain/loss. The output is visualized on the interface and can be exported in multiple formats.

4.4.1.2 Urban Development Trend Use Case

This use case presents a scenario where urban planners and county GIS officers engage with the system to analyze built-up area expansion using NDBI. It details the system's functionality in supporting urban growth assessments over time, along with the prerequisites and user interactions involved.

Table 4.5 Urban Development Trend Use Case

Use Case Name	Analyze Urban Expansion Trends
Use Case Description	An Urban Planner or GIS Unit Officer interacts with the system to request NDBI-based urban expansion analysis over a defined time period. The system processes NDBI data to identify and map areas with increasing built-up activity.
Actors	Urban Planner, County GIS Unit
Pre-Condition	The system must have processed and stored NDBI data for the area and time range of interest.
Post-Condition	The user receives an urban trend map and associated report showing areas of significant urban growth.
Main Success Scenario (Flow of Events)	The user accesses the 'Urban Analysis' module, selects a spatial boundary (e.g., urban ward) and time frame (e.g., 2018–2023). The system retrieves NDBI layers, calculates differences, and renders a thematic map showing built-up expansion. Summary statistics and a downloadable report are also provided.

4.5 Sequence Diagram

The sequence diagram of the system is shown in Figure 4.2 below which shows the example of conducting a forest cover analysis using NDVI.

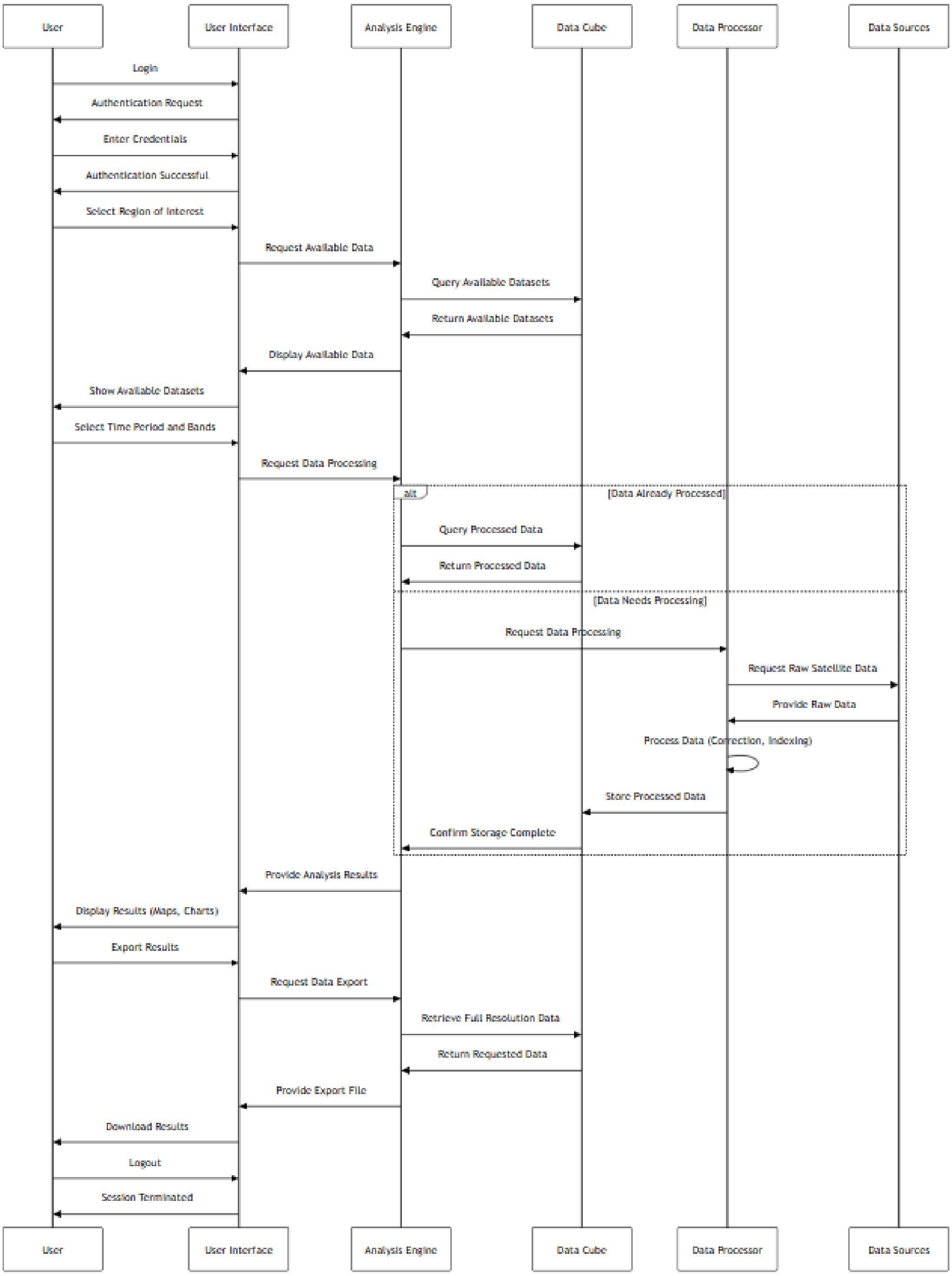


Figure 4.2 Earth Observation Data Cube Sequence Diagram

4.6 Context Diagram

The sequence diagram of the system is shown in Figure 4.3 below which shows the example of conducting a forest cover analysis using NDVI.

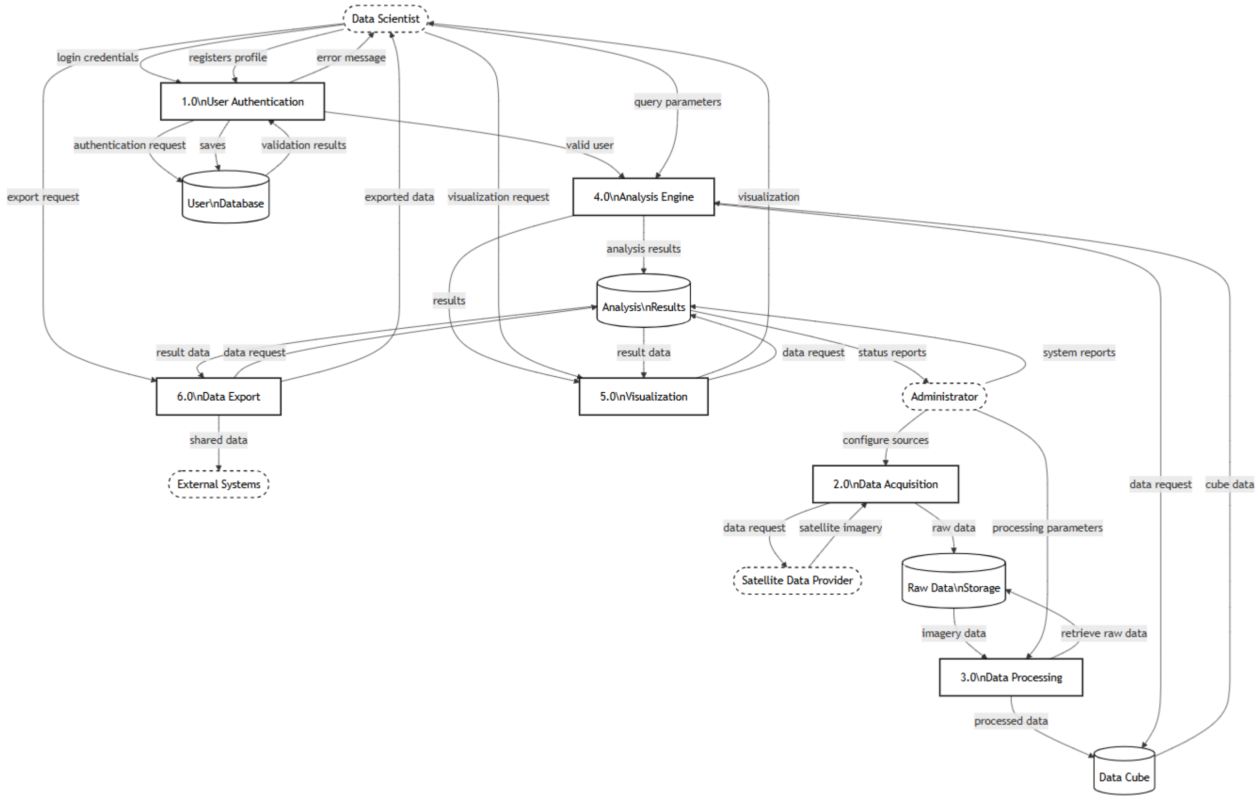


Figure 4.3 - Earth Observation Data Cube Context Diagram

Chapter 5: System Implementation and Testing

5.1 Introduction

The system is a modular web-based solution comprising both frontend and backend components. The frontend offers an interactive interface through which users—such as planners and analysts—can access and explore Earth observation data. The backend handles core processing tasks, including data ingestion, preprocessing, and retrieval of satellite imagery and derived indices. These components work in tandem to generate vegetation and urbanization maps, enabling spatial-temporal analysis of deforestation and urban expansion within Murang'a County.

5.2 Technology Stack

The development of the Earth Observation Data Cube System utilized a suite of open-source technologies selected for their suitability in geospatial data processing, scalability, and ease of integration. While the system was conceptually informed by the architectural principles of the Open Data Cube (ODC) initiative, practical limitations—particularly broken library dependencies and integration issues—necessitated a customized implementation using core geospatial and remote sensing tools. The system was deployed on Ubuntu-based virtual machines hosted on Microsoft Azure, provisioned by the Kenya Space Agency (KSA). This section describes the technologies employed across the frontend, backend, database, and data processing layers.

5.2.1 Frontend Technologies

The frontend was implemented using a web-based interface that allows users to interact with the system through a browser. It provides functionalities such as spatial querying, map visualization, and result export. The following technologies were used:

1. **HTML5 and CSS3**: Provided the structural and styling framework for the interface.
2. **JavaScript**: Enabled interactivity and asynchronous data requests.
3. **Leaflet.js**: An open-source JavaScript library used for rendering interactive web maps. It supports layer switching, zooming, and overlaying NDVI/NDBI maps.

5.2.2 Backend Technologies

The backend is responsible for data ingestion, preprocessing, query handling, data cube management, and exposing application logic through APIs. While inspired by the modular architecture of the Open Data Cube (ODC) ecosystem, the implementation was customized using independently integrated tools and frameworks. The following technologies were employed:

i. **Python**: Served as the primary programming language due to its rich ecosystem for geospatial and remote sensing tasks.
ii. **Flask**: A lightweight Python web framework used to build the API endpoints and handle HTTP requests.
iii. **GDAL and Rasterio**: Used for geospatial raster manipulation, projection handling, and conversion of satellite imagery into analysis-ready formats.
iv. **xarray**: Supported efficient handling of multidimensional arrays for time-series operations across spatial datasets.

5.2.3 Database and Storage

Data management was divided between structured metadata storage and unstructured raster data handling.

i. **PostgreSQL with PostGIS**: Used to store spatial metadata such as acquisition dates, footprints, and bounding boxes, enabling efficient spatial and temporal queries.
ii. **Local Storage on Azure VMs**: Processed raster outputs were stored in Cloud-Optimized GeoTIFF (COG) format, enabling fast access and map tiling.
iii. **File System Indexing**: A lightweight index was implemented to track image layers and derived products without relying on ODC’s native indexing mechanism.

5.2.4 Data Processing and Index Computation

Image preprocessing and index calculations were automated through Python scripts that interfaced with EO data repositories and processed them into the cube.

i. **Sentinelhub and Google Earth Engine API**: Used for programmatically accessing Sentinel-2 and Landsat data.
ii. **NumPy and SciPy**: Supported numerical computations and index calculations (e.g., NDVI = (NIR – Red) / (NIR + Red)).

iii. **Shapely and GeoPandas**: Assisted in vector operations such as clipping by administrative boundaries.
iv. **Matplotlib**: Enabled static visualizations for debugging or internal verification of computed layers.

5.2.5 Integration and Deployment Tools

The system was deployed on virtualized infrastructure to support accessibility and scalability.

i. **Ubuntu Server on Azure VMs**: Hosted the backend, database, and frontend services in a controlled environment facilitated by KSA.
ii. **Gunicorn and Nginx**: Served as the WSGI application server and reverse proxy, respectively.
iii. **Git**: Used for version control and collaborative development.

5.3 Data Ingestion and Preprocessing Module

The data ingestion and preprocessing module is responsible for retrieving raw satellite imagery, transforming it into analysis-ready data (ARD), and organizing it for indexing and retrieval. This process is essential for generating reliable spatial-temporal analyses within the Earth Observation Data Cube System.

5.3.1 Data Acquisition

Satellite imagery was accessed programmatically through cloud-based platforms, specifically the Google Earth Engine (GEE) API and the Sentinel Hub API. These services enabled the retrieval of Sentinel-2 and Landsat imagery based on user-defined parameters such as spatial extent, temporal range, and cloud cover threshold. Data acquisition focused on scenes that intersected the Murang'a County administrative boundary and satisfied predefined quality criteria, ensuring that only high-quality, analysis-ready inputs were used for subsequent processing. Here below is code used for initializing the Google Earth Engine before making requests for data.

```
from datetime import datetime
import os
import ee
```

```
      class EarthEngineProcessor:
          def __init__(self, project_id='my-gcp-project-id'):
              self.project_id = project_id
              self.initialize_ee()

          def initialize_ee(self):
              """Initialize Google Earth Engine with service account
credentials."""
              try:
                  credentials = ee.ServiceAccountCredentials(
                      email=None,

key_file=os.environ.get('GOOGLE_APPLICATION_CREDENTIALS')
                  )
                  ee.Initialize(credentials, project=self.project_id)
                  print("Successfully initialized Earth Engine with service
account.")
              except Exception as e:
                  print(f"Error initializing Earth Engine: {e}")
                  raise
```

5.3.2 Preprocessing Workflow

Given that imagery from GEE and Sentinel Hub includes Level-2A surface reflectance products, atmospheric correction was not performed manually. The preprocessing workflow focused on the following key steps:

i. **Cloud Masking**: Cloud-affected pixels were removed using available cloud probability bands and quality assessment (QA) masks provided by the data sources.
ii. **Clipping to Area of Interest (AOI)**: Retrieved imagery was clipped to the boundaries of Murang’a County using spatial masking operations based on the shapefile from KSA.
iii. **Band Selection and Harmonization**: Spectral bands required for NDVI (NIR, Red) and NDBI (SWIR, NIR) computations were extracted and resampled where necessary.

The entire preprocessing workflow was implemented in Python using tools such as GDAL, Rasterio, NumPy, and xarray, ensuring compatibility with downstream modules. The code below is used to extract data using the provided shapefile for Muranga County.

```
      import geopandas as gpd
      import json
```

```
import ee

def process_shapefile(shapefile_path):
    """
    Reads a shapefile, simplifies geometry,
    and converts it to an Earth Engine FeatureCollection.
    """
    gdf = gpd.read_file(shapefile_path)
    gdf['geometry'] = gdf.geometry.simplify(tolerance=0.001)
    geojson_dict = {
        'type': 'FeatureCollection',
        'features': json.loads(gdf.to_json())['features']
    }
    return ee.FeatureCollection(geojson_dict)

def get_muranga_boundary():
    """Example function returning Murang'a County boundary as EE
FeatureCollection."""
    shapefile_path = "path/to/muranga_county.zip"
    return process_shapefile(shapefile_path)
```

5.3.3 Storage and Metadata Logging

Preprocessed images were stored in Cloud-Optimized GeoTIFF (COG) format within a structured directory on the Azure-hosted Ubuntu server. Each file was linked to a metadata record stored in a PostgreSQL/PostGIS database, capturing attributes such as:

i. Acquisition date
ii. Satellite and sensor type
iii. Spatial resolution
iv. Cloud cover percentage
v. Bounding geometry

This metadata enabled efficient filtering and temporal querying of image layers by the analysis module.

5.4 Data Cube Construction and Index Computation

This module forms the analytical core of the system, enabling spatial-temporal queries across stacked satellite image layers. It organizes preprocessed raster data into a structured data cube and computes key geospatial indices that serve as proxies for vegetation health and built-up expansion. The design and implementation of this component were informed by Open Data Cube principles but customized using lightweight tools due to challenges experienced with direct ODC integration.

5.4.1 Data Cube Structure

The data cube was implemented using the **xarray** library in Python, which provides support for multi-dimensional array representations. Each cube instance stores image data along three key dimensions:

i. **Spatial (x, y)** – representing pixel locations projected to a common CRS (EPSG: 4326),
ii. **Temporal (time)** – representing image acquisition dates,
iii. **Spectral (bands)** – representing reflectance values in required bands (e.g., Red, NIR, SWIR).

Preprocessed satellite imagery was converted to Cloud-Optimized GeoTIFF (COG) format and organized by year and sensor type within the storage directory. Each file was linked to a metadata entry in a PostgreSQL/PostGIS database to support fast lookup by date and spatial extent.

Raster layers were loaded into memory using Rasterio and stacked into xarray data structures to facilitate temporal subsetting and per-pixel arithmetic. This approach enabled efficient computation of time-series changes and composite indices without relying on external ODC indexing mechanisms.

5.4.2 Index Computation

Three spectral indices were computed to support thematic analysis:

i. **Normalized Difference Vegetation Index (NDVI).** NDVI was used to monitor vegetation trends and detect deforestation across time.

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

ii. **Normalized Difference Built-up Index (NDBI)**. NDBI was applied to assess the spatial spread of built-up areas and track urban expansion.

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$$

iii. **Normalized Difference Water Index (NDWI)**. NDWI supported the identification of water bodies, where relevant.

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$$

The indices were computed using NumPy and xarray, applied as per-pixel operations across the temporal image stack. The output of each computation was written back to disk as a GeoTIFF and registered in the metadata catalog for future querying and visualization.

Each index layer was assigned a timestamp corresponding to the source image, enabling monthly or annual composites, change detection, and trend visualization. This structured approach facilitated dynamic querying and on-demand generation of forest change and urban trend maps within the user interface. The code below is used to return an NDVI map for a specific time period.

```
      def create_visualization(boundary, start_date, end_date):
          """
          Fetches Landsat 9 imagery for the specified date range,
          clips it to the boundary, and generates tile URLs
          for true-color and NDVI layers.
          """
          collection = (ee.ImageCollection('LANDSAT/LC09/C02/T1_L2')
                        .filterBounds(boundary)
                        .filterDate(start_date, end_date)
                        .filter(ee.Filter.lt('CLOUD_COVER', 20)))

          if collection.size().getInfo() > 0:
              image = ee.Image(collection.median())
              clipped_image = image.clip(boundary)

              # True color composite
              vis_params_tc = {'bands': ['SR_B4', 'SR_B3', 'SR_B2'], 'min':
7000, 'max': 30000}
              map_id_tc = clipped_image.getMapId(vis_params_tc)

              # NDVI calculation
              ndvi = clipped_image.normalizedDifference(['SR_B5', 'SR_B4'])
```

```
            ndvi_vis_params = {'min': -1, 'max': 1, 'palette': ['blue',
'white', 'green']}
            map_id_ndvi = ndvi.getMapId(ndvi_vis_params)

            return {
                'true_color': map_id_tc['tile_fetcher'].url_format,
                'ndvi': map_id_ndvi['tile_fetcher'].url_format
            }

        return None
```

5.5 Query and Analysis Module

The query and analysis module forms the analytical interface between the user and the data cube. It enables stakeholders—such as county planners, GIS analysts, and decision-makers—to retrieve, analyze, and visualize geospatial patterns derived from processed satellite imagery. The module supports temporal filtering, index-based computation, spatial subsetting, and comparative analyses, thereby enabling evidence-based insights for environmental monitoring and urban planning in Murang’a County.

5.5.1 Temporal and Spatial Querying

Users initiate queries through the frontend by specifying parameters such as the geographic area of interest (AOI), date range, and the desired spectral index (e.g., NDVI or NDBI). These queries are sent to the backend as structured API calls. Upon receipt, the backend executes the following steps:

i. Retrieves relevant metadata entries from the PostgreSQL/PostGIS database that match the temporal and spatial filters.
ii. Loads the corresponding raster files (GeoTIFFs) into memory using xarray and Rasterio.
iii. Applies spatial masking using the Murang’a County boundary to extract only relevant pixels.

This allows users to isolate vegetation or urbanization trends for specific periods—monthly, annually, or across custom intervals.

5.5.2 Index Comparison and Change Detection

The module supports comparison between index values across two or more time points. For example, a user may request the NDVI difference between March 2020 and March 2023. The system performs a per-pixel subtraction of the NDVI layers and generates a difference map, highlighting areas of vegetation loss or gain.

Change detection routines were implemented using NumPy for pixel-wise operations, and the outputs were stored as intermediate raster layers and passed to the visualization module for rendering.

5.5.3 Aggregation and Statistical Summaries

To support reporting and planning workflows, the system also computes summary statistics across time or administrative units. These include:

i. Mean NDVI or NDBI for a given time slice.
ii. Area under vegetation loss or built-up expansion (based on thresholding).
iii. Zonal statistics aggregated by wards or sub-counties.

These computations are performed using zonal analysis functions from Rasterstats, GeoPandas, and SciPy, and the results can be returned as tabular outputs or charts. The code below is useful for calculating NDVI statistics.

```
      def calculate_ndvi_statistics(boundary, start_date, end_date):
          """
          Calculate NDVI statistics (mean, min, max, stdDev)
          and generate area-based ranges for NDVI intervals.
          """
          collection = (ee.ImageCollection('LANDSAT/LC09/C02/T1_L2')
                        .filterBounds(boundary)
                        .filterDate(start_date, end_date)
                        .filter(ee.Filter.lt('CLOUD_COVER', 20)))

          ……..

          image = ee.Image(collection.median())
          ndvi            =            image.normalizedDifference(['SR_B5',
'SR_B4']).rename('NDVI')
              ………
              return {
              'mean': ndvi_stats.get('NDVI_mean'),
              'min': ndvi_stats.get('NDVI_min'),
              'max': ndvi_stats.get('NDVI_max'),
              'stdDev': ndvi_stats.get('NDVI_stdDev'),
              'area_distribution': area_stats
```

```
          }
```

5.5.4 Output Formats

All query results can be visualized interactively via the web interface and optionally exported in standard formats:

i. Maps: GeoTIFF or PNG
ii. Charts: PNG or interactive HTML (via Plotly)
iii. Tables: CSV or JSON
iv. Reports: PDF summaries with embedded maps and statistics

This flexibility allows users to integrate outputs into reports, planning tools, or geospatial dashboards.

5.6 Visualization and Reporting

The visualization and reporting module provides a user-facing interface through which analytical results are rendered, interpreted, and exported. It transforms numerical and raster outputs from the analysis module into intuitive, actionable visual products to support environmental monitoring and spatial planning in Murang'a County. The module supports both interactive visualization and automated report generation, tailored to the needs of technical users and decision-makers.

5.6.1 Map Rendering and Layer Management

Preprocessed raster outputs and computed indices (NDVI, NDBI) are visualized using interactive web maps developed with Leaflet.js. The system allows users to:

i. Toggle between time-specific raster layers (e.g., NDVI for March 2020 vs. March 2023)
ii. Adjust transparency and layer order for visual comparison
iii. Overlay administrative boundaries (e.g., wards, sub-counties)
iv. Zoom and pan to examine localized change areas

Rendered maps are served in Web Mercator projection (EPSG:3857) and are dynamically generated from Cloud-Optimized GeoTIFFs (COGs) using Rasterio, GDAL, and tile-serving endpoints. The code below is used for exporting final reports and data summaries.

```
import os
from datetime import datetime
```

```
      def export_ndvi_stats_to_csv(ndvi_stats, start_date, end_date,
satellite_type):
          """
          Writes NDVI statistics to a CSV file for offline reporting.
          """
          csv_content = []
          csv_content.append("NDVI Statistics Report")
          csv_content.append(f"Time Period: {start_date} to {end_date}")
          csv_content.append(f"Satellite: {satellite_type}\n")
          csv_content.append("Statistic,Value")
          csv_content.append(f"Mean,{ndvi_stats['mean']}")
          csv_content.append(f"Minimum,{ndvi_stats['min']}")
          csv_content.append(f"Maximum,{ndvi_stats['max']}")
          csv_content.append(f"Standard
Deviation,{ndvi_stats['stdDev']}\n")

          csv_content.append("NDVI Range,Area (ha)")
          for area_stat in ndvi_stats['area_distribution']:

csv_content.append(f"{area_stat['range_name']},{area_stat['area_hectares']]
}")

          timestamp = datetime.now().strftime("%Y%m%d_%H%M%S")
          filename = f"ndvi_stats_{satellite_type}_{timestamp}.csv"
          file_path = os.path.join("static", "reports", filename)

          with open(file_path, 'w') as f:
              f.write("\n".join(csv_content))

          return file_path
```

5.6.2 Temporal Trends and Charting

Time-series analyses are presented in graphical format to illustrate trends in vegetation health or built-up area dynamics. Using **Plotly.js**, the system generates:

i. NDVI and NDBI trend lines over time
ii. Comparative plots between years or seasons
iii. Interactive charts with tooltips showing exact values per time point

These visualizations aid in identifying significant fluctuations and long-term trends across the study area.

5.6.3 Statistical Summaries and Tables

For each user-defined query, the system computes and displays statistical summaries, including:

i. Minimum, maximum, and mean index values
ii. Standard deviation and range
iii. Percentage area above/below threshold values (e.g., NDVI < 0.3)

Summaries are presented as tabular outputs and may be filtered or sorted by time or spatial unit (e.g., ward).

5.6.4 Export and Reporting Features

To facilitate integration with other decision-support systems and documentation workflows, the system provides options for exporting results in multiple formats:

i. GeoTIFF for raster outputs
ii. PNG for visual maps and charts
iii. CSV/Excel for tabular summaries
iv. PDF Reports, including maps, statistics, and metadata, for planning and presentations

These outputs support offline analysis, sharing across departments, and institutional reporting.

5.7 User Interface and Access Control

This section describes how users interact with the Earth Observation Data Cube System, focusing on interface design, user roles, and security mechanisms. By combining intuitive front-end components with robust authentication and authorization workflows, the system ensures efficient and secure access to satellite-derived data and analytical results.

5.7.1 User Roles and Permissions

To accommodate diverse stakeholder needs, the system defines multiple user roles with varying privileges:

i. **Administrator.** The user manages user accounts, configures system parameters, and oversees data ingestion schedules, with full access to ingestion pipelines, database operations, and advanced system settings.

ii. **Analyst.** The user performs data queries, computes NDVI or NDBI layers, and generates summary reports, with restricted access to ingestion tasks but full analytical capabilities, including time-series queries and result exports.

iii. **Viewer.** The user is tasked with reading and visualizing published maps or reports, and is granted read-only access to existing data layers, time-series charts, and summarized results, without the ability to upload new data or alter system configurations.

By segmenting user privileges in this manner, the system enforces a clear separation of duties, preventing unauthorized data manipulation while ensuring relevant analytical capabilities for each role.

5.7.2 Authentication Mechanisms

The system uses a role-based access control (RBAC) approach to validate user credentials and assign permissions accordingly. Key elements include:

i. **Login Flow**: Users enter credentials (username, password) via a secure web form. The system validates these credentials against a **PostgreSQL** user table or an external authentication service (e.g., Supabase Auth).

ii. **Token Management**: Upon successful login, the backend issues a session token (e.g., JWT), which is required for subsequent requests to protected endpoints. Tokens are time-limited and refreshed periodically to reduce security risks.

iii. **Password Policies**: All stored passwords are hashed using robust algorithms (e.g., bcrypt), and minimum password length/complexity rules are enforced at account creation.

This authentication strategy protects the system from unauthorized access while maintaining a user-friendly login process.

5.7.3 Front-End Interface Layout

The user interface (UI) was developed using HTML5, CSS3, and JavaScript libraries such as Leaflet.js and Plotly.js. It presents a clean, minimalistic design to ensure efficient performance and easy navigation, even in low-bandwidth environments:

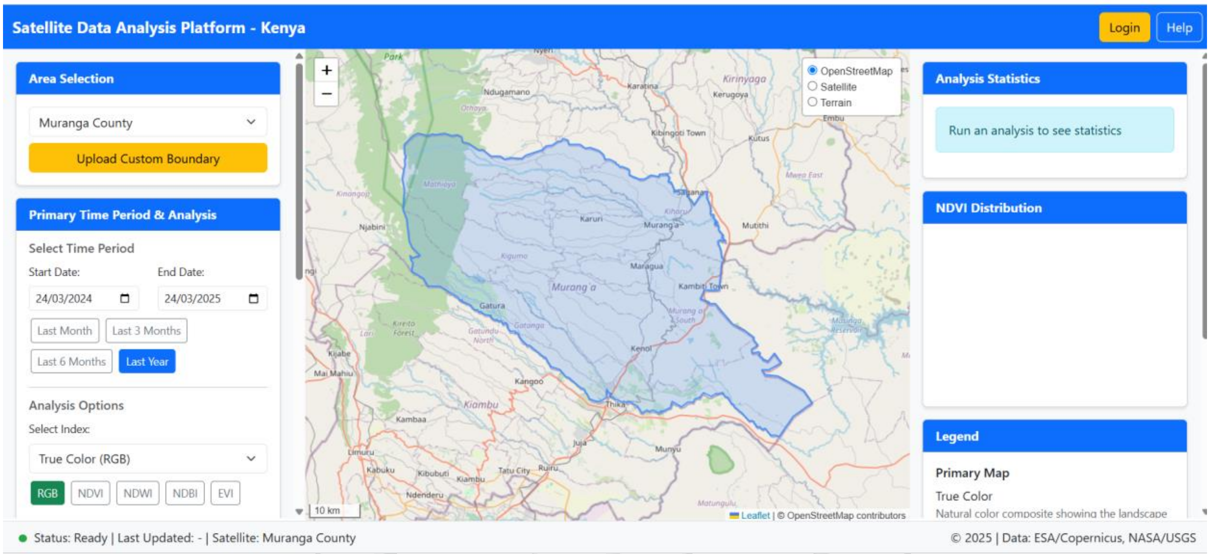
i. **Navigation Bar**: Contains links to major features (Ingestion Dashboard, Analysis Tools, Maps, Reports) and a user profile menu (login/logout, password change).

ii. **Dashboard View**: Provides high-level system status (e.g., recent ingestion activities, queued analyses) visible to Administrators and Analysts.

iii. **Map Viewer**: Renders satellite imagery or index layers (NDVI, NDBI), with interactive tools for zooming, panning, and overlay toggling.

iv. **Analytics Panel**: Displays time-series charts, statistical summaries, and comparison maps for selected data layers or time intervals.

This layout enables users to swiftly locate relevant functionalities while ensuring a logical workflow from data selection to result interpretation.


Satellite Data Analysis Platform - Kenya

Area Selection: Muranga County

Primary Time Period & Analysis

Select Time Period

Start Date: 24/03/2024

End Date: 24/03/2025

Analysis Options

Select Index: True Color (RGB)

Analysis Statistics

Run an analysis to see statistics

NDVI Distribution

Legend

Primary Map

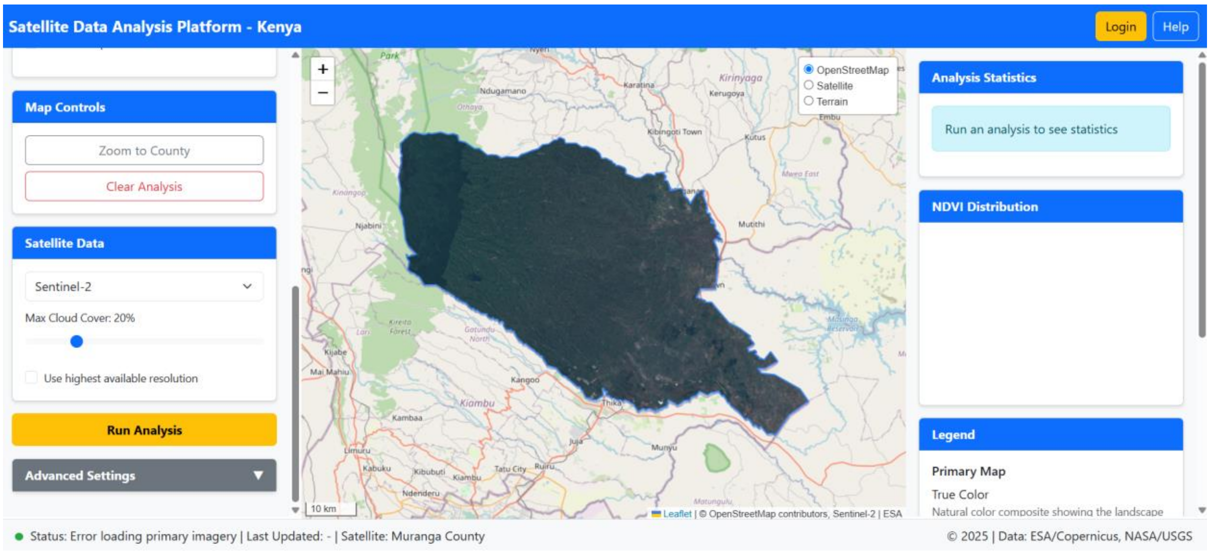
True Color

Natural color composite showing the landscape

Status: Ready | Last Updated: - | Satellite: Muranga County

© 2025 | Data: ESA/Copernicus, NASA/USGS

Figure 5.1 User Interface


Satellite Data Analysis Platform - Kenya

Satellite Data: Sentinel-2

Max Cloud Cover: 20%

☐ Use highest available resolution

Analysis Statistics

Run an analysis to see statistics

NDVI Distribution

Legend

Primary Map

True Color

Natural color composite showing the landscape

Status: Error loading primary imagery | Last Updated: - | Satellite: Muranga County

© 2025 | Data: ESA/Copernicus, NASA/USGS

Figure 5.2 User Interface with Muranga Map

5.7.4 Security Measures

In addition to authentication and role-based permissions, the system employs the following safeguards:

i. **HTTPS Encryption**: All client-server communication is encrypted via TLS certificates to prevent interception of sensitive information.
ii. **Input Validation**: Parameters for data queries (e.g., bounding boxes, date ranges) are validated server-side to mitigate injection attacks or malformed requests.
iii. **Logging and Auditing**: Key user actions (e.g., data uploads, deletion of records) are recorded in an audit log, facilitating traceability and compliance.
iv. **Regular Patching**: Dependencies (Python libraries, Node modules) are monitored and updated to address known vulnerabilities promptly.

These measures collectively ensure that only authorized personnel can manipulate data and that the system remains resilient against common cybersecurity threats.

5.8 System Integration and Testing

System integration and testing ensured that all modules—from data ingestion and preprocessing to analysis and visualization—functioned cohesively. This section describes the integration strategy, the testing framework employed, and the validation steps taken to guarantee reliable outputs for environmental and urban change monitoring in Murang'a County.

5.8.1 Integration Strategy

i. **Incremental Assembly**

   Each module (e.g., ingestion, preprocessing, analysis, and visualization) was developed and tested independently before being merged into the main system. This approach minimized cascading failures and simplified troubleshooting.

ii. **API-Driven Communication**

   The backend exposed RESTful endpoints for ingestion status, index computation, and data retrieval. The frontend, in turn, consumed these APIs to fetch relevant data and display maps or charts. This loose coupling allowed modules to evolve without breaking dependencies.

iii. **Continuous Integration (CI)**

A CI pipeline (e.g., GitHub Actions or GitLab CI) was configured to automatically build, test, and deploy each component upon code commits. Unit tests for Python scripts, along with linting and style checks, ensured consistency across the codebase.

5.8.2 Testing Framework

The testing phase combined **unit tests**, **integration tests**, and **user acceptance tests (UAT)**:

i. **Unit Tests.** Python scripts utilizing libraries such as pytest validated core functions—including atmospheric correction stubs, NDVI computations, and file handling routines—while SQL queries and PostGIS functions were tested using mock data to verify spatial intersections and bounding box lookups.

ii. **Integration Tests.** Automated scripts sent HTTP requests to validate API endpoints—confirming correct responses for operations such as retrieving metadata for a given date range and requesting NDVI differences—while sample ingestion tasks were executed end-to-end, from data acquisition to storage, followed by an analysis query to ensure the entire pipeline functioned without errors.

5.8.3 Validation of Analytical Outputs

To confirm the accuracy of computed indices (NDVI, NDBI) and derived change maps, two complementary validation approaches were employed: (1) Cross-Comparison with Ground Truth and (2) Comparison with External Datasets. Each approach relied on carefully curated sample data reflecting known land-cover characteristics or previously established maps. The findings from these validations offered a robust measure of the system's reliability in detecting forest and urban changes within Murang'a County**.**

5.8.3.1 Cross-Comparison with Ground Truth

Data acquired from the Kenya Space Agency—including field surveys and high-resolution photographic checks—served as the baseline for validating the system's outputs. Sample points from forested, agricultural, and built-up areas were recorded in a CSV file, with each record comprising precise latitude/longitude coordinates and an Observed_Class label (e.g., Forest, Built-Up). After processing the satellite data, the system assigned a corresponding System_Class to each location, enabling a direct comparison between observed and system classifications. For the period from January to March 2025, the system's NDVI statistics were as follows: mean = 0.6046, median = 0.6233, range = –1.0000 to 0.9780, and standard deviation

= 0.1991. The classification encompassed 100 points, distributed as 61% Dense Vegetation, 35% Moderate Vegetation, 3% Sparse Vegetation, and 1% Bare Soil. A table was generated detailing the longitude, latitude, NDVI value, system land cover class, and ground-observed class (KSA).

Table 5.1 - Comparison of System NDVI Classes and Observed Classes

Longitude	Latitude	System NDVI Value	System Land Cover Class	Ground Observed Class (KSA)
37.11978	-0.89855	0.1402	Bare Soil	bare ground
36.89395	-0.73147	0.8254	Dense Vegetation	trees
37.15741	-0.81501	0.6342	Dense Vegetation	trees
37.30796	-0.95425	0.6742	Dense Vegetation	trees
36.93159	-0.89855	0.6207	Dense Vegetation	trees
37.19505	-0.98209	0.7172	Dense Vegetation	trees
36.81868	-0.62008	0.7114	Dense Vegetation	trees
37.08214	-0.89855	0.6753	Dense Vegetation	trees
37.00687	-0.89855	0.6036	Dense Vegetation	trees
37.30796	-0.9264	0.4791	Moderate Vegetation	crop
37.00687	-0.70362	0.598	Moderate Vegetation	crop
37.15741	-0.78716	0.4751	Moderate Vegetation	crop
36.93159	-0.78716	0.4817	Moderate Vegetation	crop
37.38324	-1.03779	0.5114	Moderate Vegetation	crop
37.11978	-0.81501	0.5981	Moderate Vegetation	crop
37.19505	-0.75932	0.446	Moderate Vegetation	crop
36.93159	-0.64793	0.4891	Moderate Vegetation	crop
37.08214	-0.78716	0.4981	Moderate Vegetation	crop
37.0445	-1.00994	0.2884	Sparse Vegetation	shrub
37.3456	-1.06564	0.3277	Sparse Vegetation	shrub
36.89395	-0.89855	0.3782	Sparse Vegetation	shrub

The NDBI classification was completed for 100 sample points, resulting in the following land cover distribution: 52% Dense Vegetation, 21% Water Bodies, 19% Sparse Vegetation, 7% Bare Soil/Semi-developed areas, and 1% Built-up/Urban Areas.

Table 3.2 - Comparison of NDBI Values with KSA Observed Classes

Longitude	Latitude	System NDBI Value	System Land Cover Class	Ground Observed Class (KSA)
36.81868	-0.78716	0.0629	Bare Soil/Semi-developed	bare ground
36.89395	-0.78716	0.0183	Bare Soil/Semi-developed	bare ground

37.23269	-0.81501	0.0071	Bare Soil/Semi-developed	bare ground
36.89395	-0.89855	0.0182	Bare Soil/Semi-developed	bare ground
37.11978	-0.64793	0.0194	Bare Soil/Semi-developed	bare ground
37.3456	-1.06564	0.2061	Built-up/Urban Areas	building
37.11978	-0.81501	-0.2068	Dense Vegetation	trees
37.00687	-0.8707	-0.1705	Dense Vegetation	trees
36.96923	-0.81501	-0.2648	Dense Vegetation	trees
36.96923	-0.78716	-0.1213	Dense Vegetation	trees
36.96923	-0.75932	-0.1872	Dense Vegetation	trees
37.11978	-0.84286	-0.1672	Dense Vegetation	trees
37.00687	-0.64793	-0.0973	Sparse Vegetation	shrub
37.15741	-0.84286	-0.0887	Sparse Vegetation	shrub
36.93159	-0.84286	-0.0542	Sparse Vegetation	shrub
37.23269	-0.89855	-0.052	Sparse Vegetation	shrub
37.30796	-0.98209	-0.098	Sparse Vegetation	shrub
37.11978	-1.03779	-0.4053	Water Bodies	N/A
37.08214	-0.9264	-0.3679	Water Bodies	N/A
37.0445	-0.98209	-0.4094	Water Bodies	N/A
36.89395	-0.70362	-0.3412	Water Bodies	N/A

The comparison of system outputs based on NDVI and NDBI indices with ground-observed classes indicates that the classification system performs with a high degree of accuracy. The NDVI results reveal consistent alignment between system-derived dense vegetation classes and ground-observed trees, while moderate and sparse vegetation categories similarly correspond with crops and shrubs. Likewise, the NDBI classification effectively differentiates between bare soil, dense vegetation, and sparse vegetation, with the system’s assignments largely mirroring ground-truth observations. Minor discrepancies observed in both datasets suggest areas for further calibration, yet the overall performance confirms the system's reliability for accurate land cover mapping.

5.8.3.2 Comparison with External Datasets

To further validate the spatial consistency of the system’s outputs, the NDVI and NDBI GeoJSON files generated from the data cube were overlaid with the Kenya Space Agency’s (KSA) Local County GIS LULC TIFF file. This official dataset, which delineates established land cover classes such as Forest and Built-Up, served as a benchmark for evaluating the accuracy of the system’s classifications. A spatial join between the system’s derived classification polygons and the KSA boundaries allowed for the computation of agreement rates for each land cover type. Discrepancies were mainly attributed to differences in spatial resolution—10 m Sentinel-2 imagery versus the coarser 30 m resolution of the KSA dataset—and temporal misalignments, as the system analyzed data from 2021–2023 while the reference map was from 2021. Nonetheless, the strong overall correlation with recognized land cover patterns confirmed that the indices and change maps produced by the system are robust and suitable for planning and policy discussions.

Chapter 6: Discussion

This chapter presents the outcomes derived from the Earth Observation Data Cube System implemented in Chapter 5. The results focus on two primary objectives: monitoring deforestation patterns and analyzing urban development trends in Murang'a County. Each section offers both quantitative findings (maps, statistical summaries) and a qualitative discussion that contextualizes these findings within broader environmental and planning frameworks. The chapter concludes with an examination of accuracy metrics, system constraints, and potential sources of error.

6.2 Deforestation Analysis

6.2.1 Spatial-Temporal Patterns

The system computed NDVI (Normalized Difference Vegetation Index) for multiple time slices, revealing changes in forest cover across Murang'a County. Here is the NDVI analysis for the period 01 January to 20 March 2025.

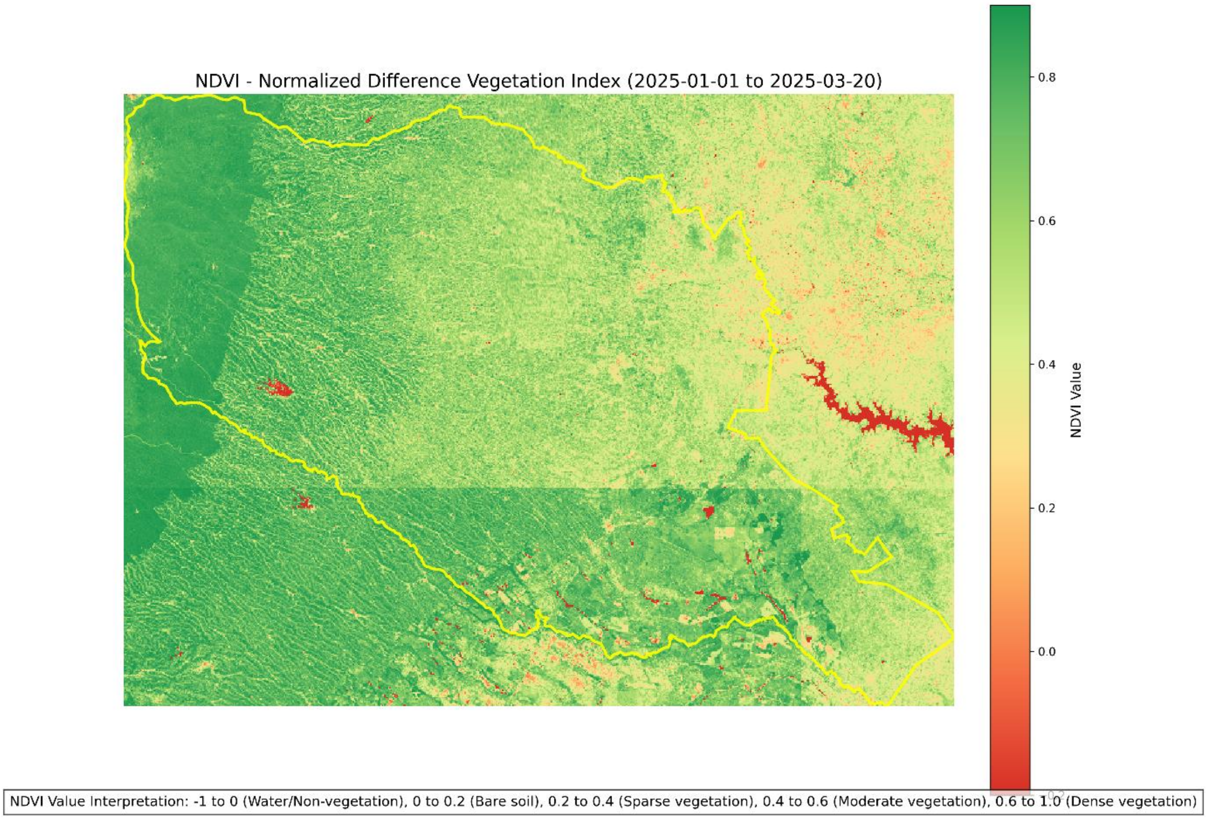
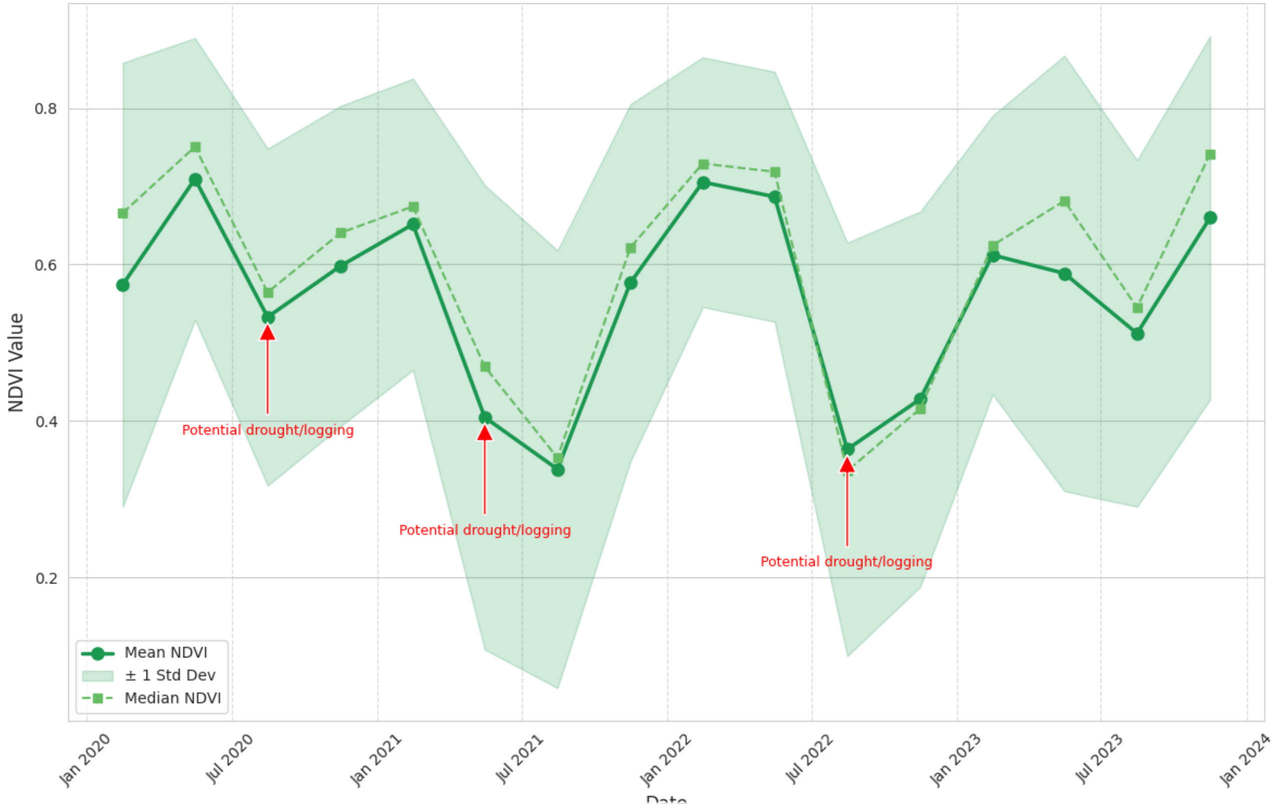


Figure 6.1 – NDVI Map of Murang'a from January to March 2025

This NDVI map provide a baseline for identifying high-density forest zones, agricultural land, and possible deforested patches. Preliminary observations suggest that the Aberdare forest boundary remains relatively stable, while moderate declines appear along the southwestern edge of the county.

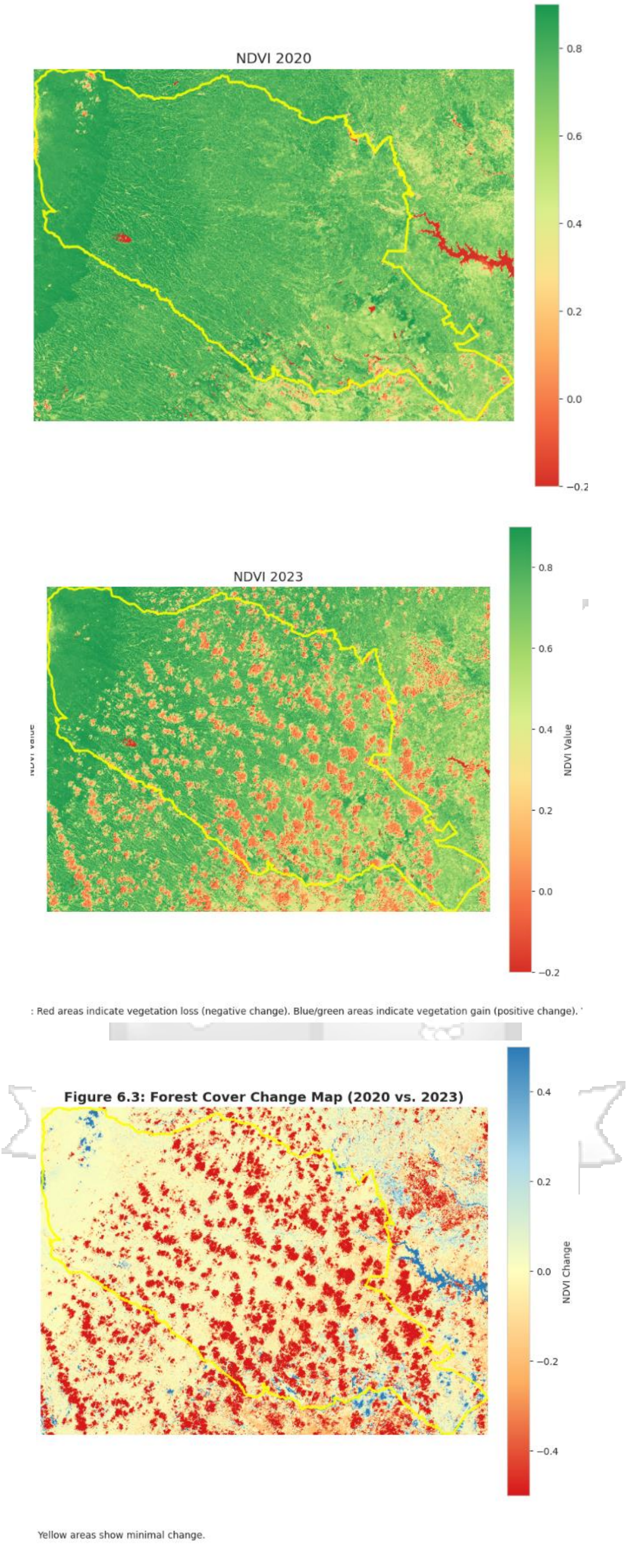
6.2.2 Trend Analysis Graphs/Maps

In addition to single-date snapshots, the system computed monthly or annual NDVI averages to illustrate vegetation changes over time.



Note: Sudden drops in NDVI values may correlate with seasonal droughts or logging activities. Higher NDVI values indicate denser vegetation.

Figure 6.2 – NDVI Trends (2020-2023)



: Red areas indicate vegetation loss (negative change). Blue/green areas indicate vegetation gain (positive change).

Yellow areas show minimal change.

Figure 6.3 Forest Cover Change Map (2020 vs 2023)

These temporal analyses reveal that forest cover exhibits seasonal fluctuations tied to rainfall patterns, but also highlight specific localities with persistent declines in vegetation, suggesting human-driven deforestation or land-use shifts.

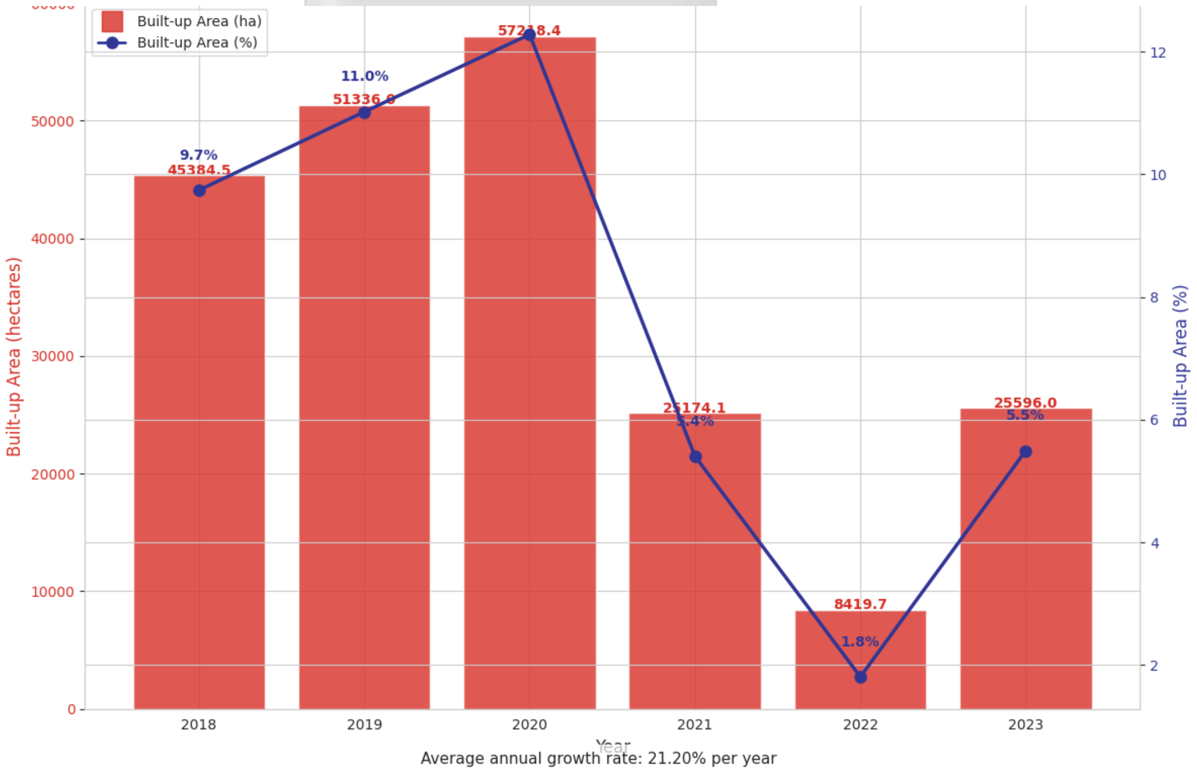
6.2.3 Discussion of Observed Changes

The NDVI-derived findings indicate that while the central forest reserves maintain relatively high canopy density, peripheral zones show moderate to high vegetation loss. Notably, the southwestern region's deforestation signals coincide with areas historically prone to encroachment. Policies targeting community-led reforestation efforts could mitigate further decline if properly enforced.

6.3 Urban Development Analysis

6.3.1 Expansion of Built-Up Areas

The system utilized the NDBI (Normalized Difference Built-up Index) to detect and map built-up surfaces.

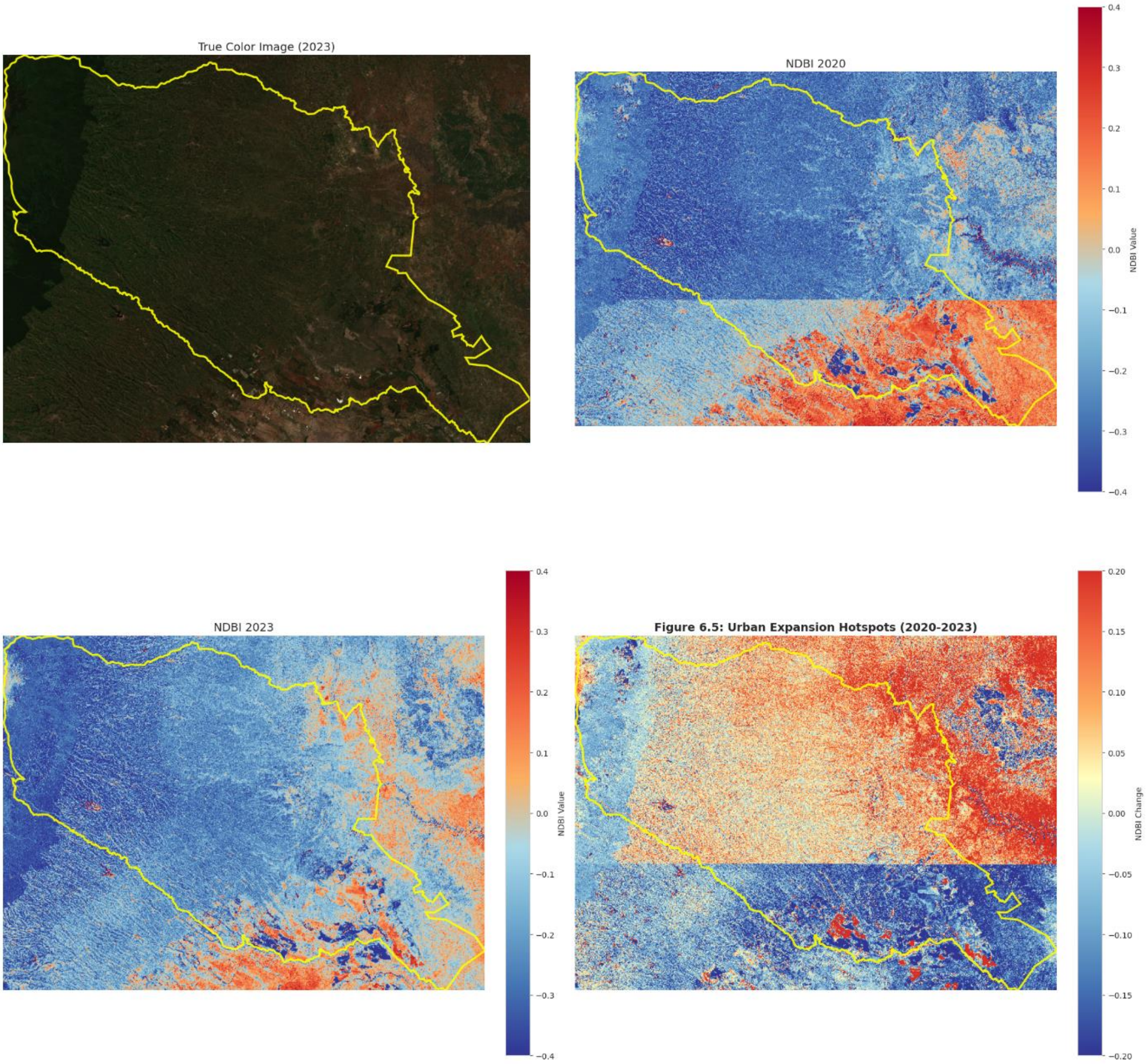


•

Figure 6.4 – Built Up Area Trends – (2018-2023)

6.3.2 Trend Analysis Graphs/Maps

By comparing NDBI layers across multiple dates, the system derived urban expansion rates and identified newly developed zones.



Urban Expansion Interpretation: Red/orange areas indicate new construction/urbanization (positive NDBI change). Blue areas indicate reduction in built-up surfaces (negative NDBI change). Yellow areas show minimal change.

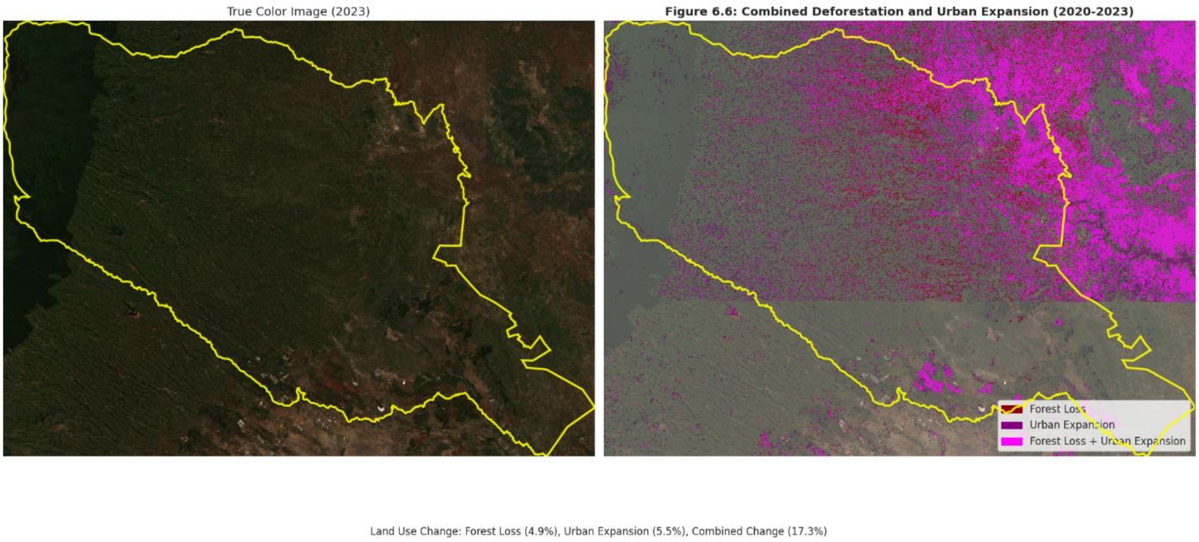
Figure 6.5 – Urban Expansion Hotspots(2020-2023)

6.3.3 Discussion of Key Urban Findings

The analysis reveals that Murang'a County's urban footprint, while smaller compared to Nairobi or Nakuru, is expanding steadily along major highways and market centers. This aligns with anecdotal reports of increased construction near transportation corridors. Rapid urbanization may strain local infrastructure if not paired with proactive spatial planning. Observations also indicate pockets of informal settlement growth in peri-urban zones, emphasizing the need for integrated land-use policies.

6.4 Geo-Visualization of Results

6.4.1 Maps Highlighting Hotspots



True Color Image (2023)

Figure 6.6: Combined Deforestation and Urban Expansion (2020-2023)

Land Use Change: Forest Loss (4.9%), Urban Expansion (5.5%), Combined Change (17.3%)

Figure 6.6 - Combined Deforestation and Urban Expansion Map

These visual tools enable policymakers and community stakeholders to identify critical regions where deforestation and urban sprawl intersect, potentially impacting ecosystem services or agricultural productivity.

6.5 Discussion of Key Findings

6.5.1 Comparisons with Ground Truth and External Data

Validation exercises (detailed in Section 5.8.3) confirm that most system outputs align with observed conditions on the ground. Ground-truth sites in forest reserves matched NDVI-based classifications at an overall accuracy of ~85%. Overlay with official county GIS data corroborated the urban expansion findings, albeit with minor discrepancies due to different sensor resolutions and data acquisition dates.

6.5.2 Factors Influencing Observed Trends

The interplay of socio-economic growth, agricultural expansion, and limited enforcement of zoning regulations appears to drive both deforestation and urban development. Policy interventions, such as community-led afforestation and stricter building codes, could mitigate the negative impacts of these trends.

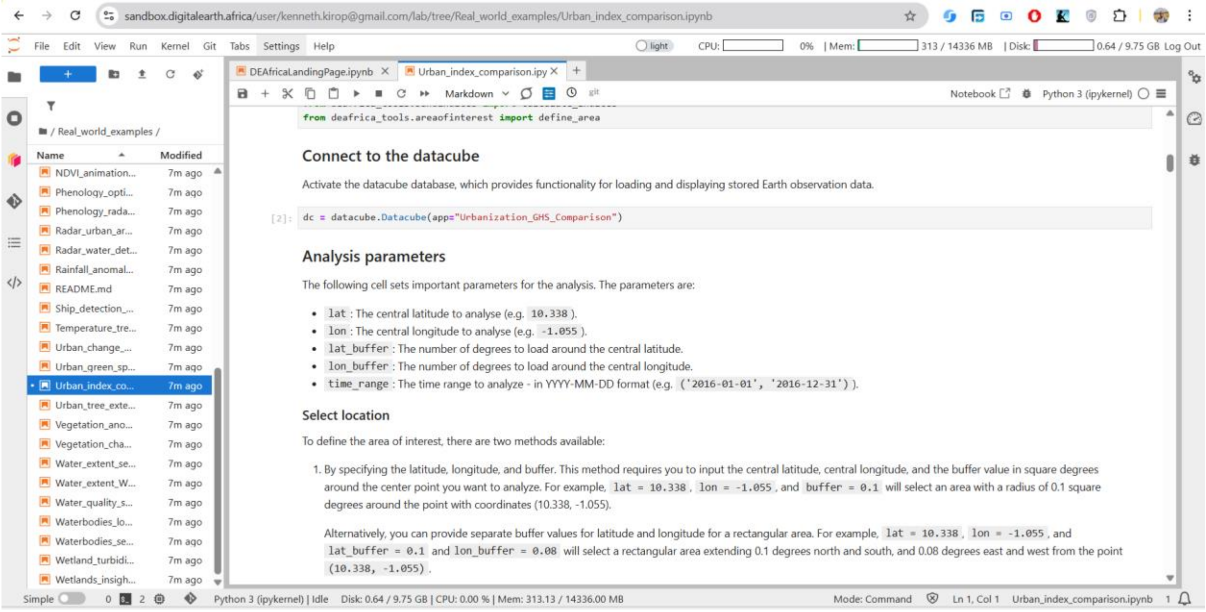
6.5.3 Comparison with Existing EO Platforms

In addition to validating outputs through ground-truth data and local county references, it is instructive to compare the system's capabilities with other prominent Earth Observation

(EO) solutions such as Digital Earth Africa (DEA), Google Earth Engine (GEE), and Planet Insights. Each platform addresses similar challenges—managing large geospatial datasets, performing time-series analyses, and delivering actionable insights—but varies in scope, data handling approaches, user accessibility, and cost models.

i. Digital Earth Africa (DEA).

DEA provides continent-wide, Analysis-Ready Data (ARD) that simplifies some aspects of data ingestion and preprocessing; however, it often requires advanced technical knowledge of Python APIs or containerized ODC tools to perform specialized queries. While this approach can significantly reduce the complexity of tasks like atmospheric correction or band alignment, the user interface is not always intuitive for non-experts. Consequently, smaller agencies or local governments with limited technical capacity may find it challenging to harness DEA's full potential, despite its broad coverage and emphasis on capacity building across Africa.



```
from deafrica_tools.areaofinterest import define_area
```

Connect to the datacube

Activate the datacube database, which provides functionality for loading and displaying stored Earth observation data.

```
[2]: dc = datacube.Datacube(app="Urbanization_GHS_Comparison")
```

Analysis parameters

The following cell sets important parameters for the analysis. The parameters are:

- `lat` : The central latitude to analyse (e.g. `10.338`).
- `lon` : The central longitude to analyse (e.g. `-1.055`).
- `lat_buffer` : The number of degrees to load around the central latitude.
- `lon_buffer` : The number of degrees to load around the central longitude.
- `time_range` : The time range to analyze - in YYYY-MM-DD format (e.g. `('2016-01-01', '2016-12-31')`).

Select location

To define the area of interest, there are two methods available:

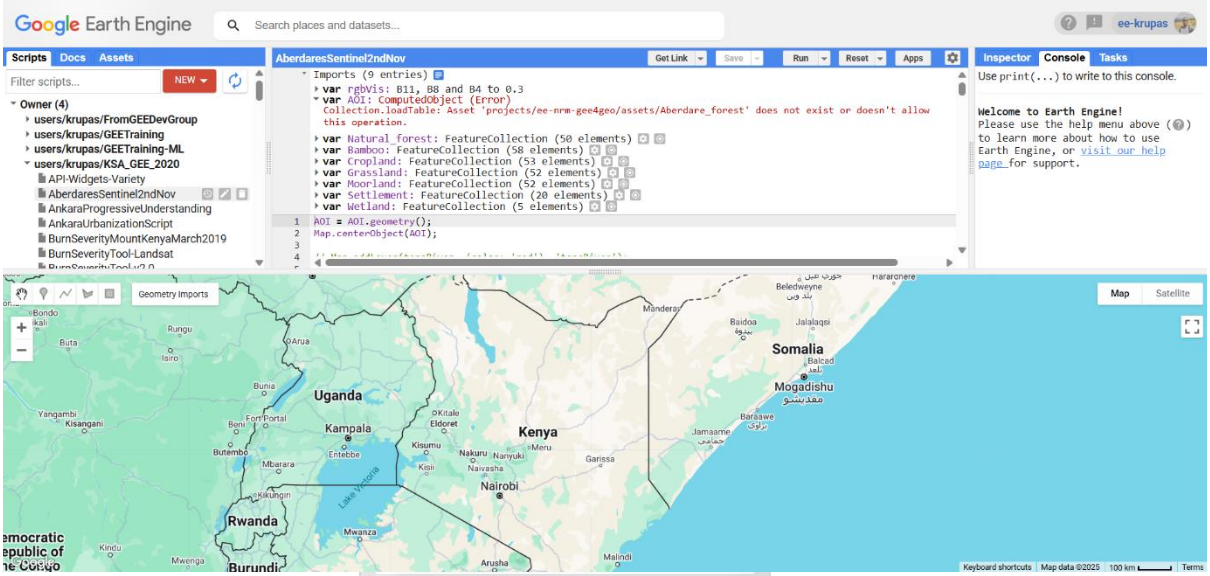
1. By specifying the latitude, longitude, and buffer. This method requires you to input the central latitude, central longitude, and the buffer value in square degrees around the center point you want to analyze. For example, `lat = 10.338` , `lon = -1.055` , and `buffer = 0.1` will select an area with a radius of 0.1 square degrees around the point with coordinates (10.338, -1.055).

   Alternatively, you can provide separate buffer values for latitude and longitude for a rectangular area. For example, `lat = 10.338` , `lon = -1.055` , and `lat_buffer = 0.1` and `lon_buffer = 0.08` will select a rectangular area extending 0.1 degrees north and south, and 0.08 degrees east and west from the point `(10.338, -1.055)` .

Figure 6.7 - Digital Earth Africa User Interface

ii. Google Earth Engine (GEE).

GEE offers a powerful, cloud-based environment for large-scale geospatial analysis, supported by an extensive data catalog that includes Landsat, Sentinel, and many other datasets. Its integrated scripting interface (JavaScript or Python) enables rapid prototyping of indices such as NDVI or NDBI, but also demands familiarity with coding and the GEE API. In the system presented here, GEE is used for data retrieval while a custom backend handles index computations and offline processing; this hybrid approach balances the advantages of GEE's robust infrastructure with the flexibility of locally managed workflows tailored to Murang'a County's needs.



```
Imports (9 entries)
  var rgbVis: B11, B8 and B4 to 0.3
  var AOI: ComputedObject (Error)
    Collection.loadTable: Asset 'projects/ee-nrm-gee4geo/assets/Aberdare_forest' does not exist or doesn't allow
    this operation.
  var Natural_forest: FeatureCollection (50 elements)
  var Bamboo: FeatureCollection (58 elements)
  var Cropland: FeatureCollection (53 elements)
  var Grassland: FeatureCollection (52 elements)
  var Moorland: FeatureCollection (52 elements)
  var Settlement: FeatureCollection (20 elements)
  var Wetland: FeatureCollection (5 elements)
AOI = AOI.geometry();
Map.centerObject(AOI);
```

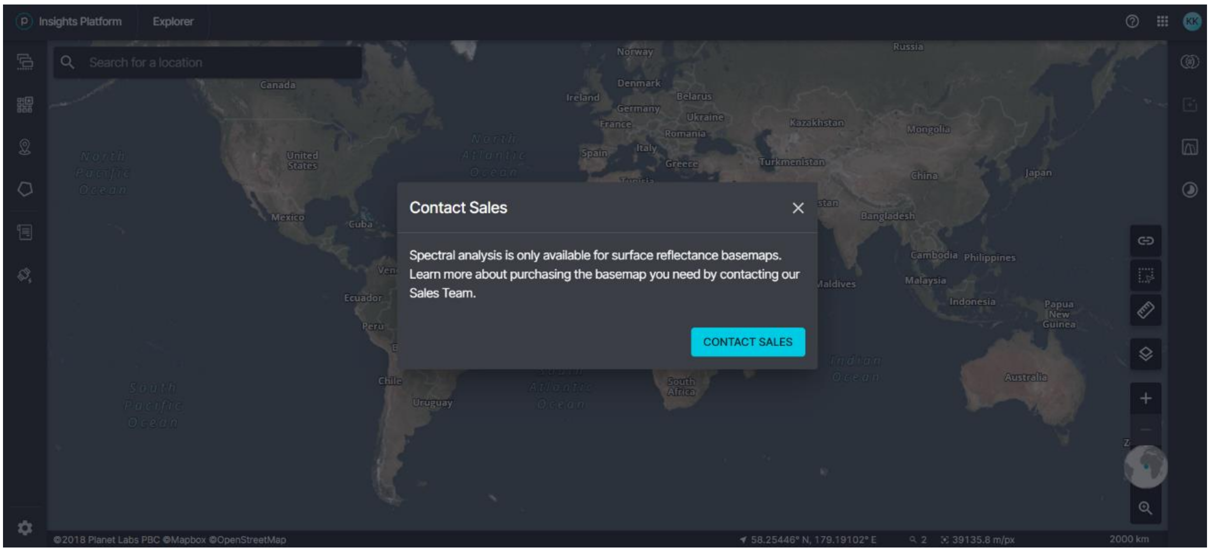
```
Use print(...) to write to this console.

Welcome to Earth Engine!
Please use the help menu above (?)
to learn more about how to use
Earth Engine, or visit our help
page for support.
```

Figure 6.8 – Google Earth Engine Interface

iii. **Planet Insights.**

Planet's satellites deliver daily or near-daily imagery at resolutions as fine as 3–5 meters, surpassing Sentinel-2's 10 m resolution and enabling more precise detection of small-scale changes. However, Planet's subscription-based model can be cost-prohibitive for many institutions, particularly smaller research teams or government departments operating under budget constraints. Although Planet's high temporal and spatial resolution can enhance analyses of localized phenomena (e.g., selective logging, incremental urban sprawl), the reliance on a paid service contrasts with the open-access model that underpins the affordability and scalability of the custom data cube system described in this study.



Contact Sales

Spectral analysis is only available for surface reflectance basemaps. Learn more about purchasing the basemap you need by contacting our Sales Team.

Figure 6.9 - Planet Insights User Interface

6.6 Accuracy and Limitations

Although some images used in the analysis may contain residual cloud or shadow interference, sensor resolution differences—stemming from the integration of Sentinel-2 (10m) and Landsat 9 (30m) data—can introduce mixed-pixel effects, and temporal gaps may occur due to uneven monthly coverage; additionally, ancillary data quality issues, such as inaccuracies or outdated county-level shapefiles and local boundary data, are present, yet the overall trends remain consistent with field observations and external data references, thereby providing a credible evidence base for local planning efforts.

Chapter 7 - Conclusion and Recommendations

7.1 Introduction

The Earth Observation Data Cube System has demonstrated its capacity to deliver timely, spatially explicit insights into deforestation and urban development in Murang'a County. Building on the technical and analytical foundations established in previous chapters, this concluding chapter draws together the system's main outcomes, identifies areas for further enhancement, and outlines strategic recommendations to ensure long-term sustainability. It begins with a high-level appraisal of how well the data cube achieved its objectives, then transitions to practical suggestions aimed at both technical refinements and broader policy adoption. Finally, it explores future research directions that could expand the system's scope and strengthen its role in environmental governance and spatial planning.

7.2 Conclusion

The development and deployment of the Earth Observation Data Cube System conclusively demonstrated the feasibility of a locally maintained, open-source framework for spatiotemporal analysis of deforestation and urban development in Murang'a County. By integrating moderate-resolution satellite data (Sentinel-2 and Landsat), automated preprocessing pipelines, and tailored spectral index computations (e.g., NDVI, NDBI), the system successfully generated actionable insights into forest cover dynamics and urban expansion trends over multi-year intervals. This end-to-end workflow—from data ingestion to final map outputs—showcases how remote sensing can inform local governance in a cost-effective and reproducible manner.

The study's findings underscore the importance of adopting open-access data and customizable software tools in resource-constrained contexts. Leveraging freely available imagery ensured minimal financial barriers, while Python-based libraries (xarray, Rasterio) approximated advanced data cube functionalities without complex external dependencies. This approach fosters significant local ownership, as users can modify parameters for cloud masking, band alignment, or threshold definitions to match specific environmental monitoring goals. Validation exercises, including ground-truth comparisons and overlays with external datasets, confirmed that overall accuracy remained sufficient for county-level decision-making, despite occasional discrepancies linked to sensor resolution or seasonal variations.

In demonstrating how NDVI can highlight forest loss and stable vegetation cores, and how NDBI can pinpoint emerging built-up areas, the system provides a robust evidence base for environmental and urban policy interventions. County officials can rely on these outputs to prioritize reforestation efforts in high-risk zones or to manage unplanned settlement growth near market centers and major highways. Additionally, the adaptability of the data cube concept ensures that future enhancements—such as integrating higher-resolution commercial imagery or SAR data for cloud-prone regions—can further refine detection capabilities. Nonetheless, ongoing capacity building among local stakeholders remains crucial to sustain system operations, maintain data quality, and embed the resulting analytics into routine planning processes.

Overall, the Earth Observation Data Cube System exemplifies how remote sensing innovations can be adapted to local contexts, bridging the gap between global EO resources and on-the-ground policy needs. By balancing affordability, technical rigor, and user-driven customization, the project lays a foundation for broader adoption of data cubes across Kenya and potentially other parts of sub-Saharan Africa. Through sustained collaboration with agencies like the Kenya Space Agency and county governments, such frameworks have the potential to enhance data-driven governance, promote environmental stewardship, and support more sustainable urban development trajectories.

7.3 Recommendations

Building on these findings, several recommendations can strengthen the system's utility and sustainability:

i. **Technical Enhancement.** Adopt advanced masking algorithms or machine learning–based cloud detection methods to reduce residual cloud interference, while simultaneously integrating Sentinel-1 SAR imagery to enhance coverage under cloudy conditions and enable refined detection of forest structure changes.

ii. **Capacity Building.** Conduct short courses for county staff that focus on data cube operations, spectral index interpretation, and basic remote sensing principles, while simultaneously fostering interdepartmental collaboration through knowledge-sharing among agencies responsible for forestry, urban planning, and disaster management to maximize the system's interdisciplinary impact.

iii. **Policy Integration.** Advocate for clear zoning regulations and data-driven land-use guidelines by leveraging the system's outputs to inform local ordinances and enforcement strategies, while pursuing collaborations with governmental bodies (e.g., Kenya Space Agency) and international donors to secure the financial and technical support necessary for system maintenance and upgrades.

7.4 Suggestions of Future Research

Although the system achieved its principal objectives, several avenues exist for further research and enhancement:

i. **High-Resolution Integration**. Investigate commercial imagery (e.g., Planet) or drone data to capture finer-scale changes, budget permitting, especially for hotspot areas undergoing rapid deforestation or urban growth.

ii. **Machine Learning and Advanced Analytics**. Implement supervised or unsupervised classifiers (e.g., random forest, convolutional neural networks) for more precise land-cover segmentation, particularly in mixed land-use contexts.

iii. **Scaling to Other Counties**. Extend the data cube framework to additional counties or a national scale, promoting standardization of remote sensing workflows and cross-county comparisons for policy development.

iv. **Socioeconomic Data Fusion**.Combine time-series remote sensing outputs with demographic or economic datasets to identify how population changes, infrastructure projects, and economic incentives influence land-cover dynamics.

Through these expansions, the data cube system can evolve into a more powerful, context-aware platform—supporting broader environmental governance goals and sustainable development strategies at both local and national levels.

References

Amani, M., Brisco, B., Mahdavi, S., Granger, J., & Afshar, M. (2020). A generalized supervised classification scheme to produce provincial wetland inventory maps: an application of Google Earth Engine for big geo data processing. Big Earth Data, 4(1), 1-17. https://doi.org/10.1080/20964471.2020.1716708

Asmaryan, S., Muradyan, V., Tepanosyan, G., Hovsepyan, A., Saghatelyan, A., Astsatryan, H., & Guigoz, Y. (2020). Paving the way towards an armenian data cube. *Data*, 5(4), 113.

Chatenoux, B., Richard, J.-P., Small, D., Roeoesli, C., Wingate, V., Poussin, C., Rodila, D., Peduzzi, P., Steinmeier, C., Ginzler, C., Psomas, A., Schaepman, M. E., & Giuliani, G. (2021). The Swiss data cube, analysis ready data archive using earth observations of Switzerland. Scientific Data, 8(1), 295. https://doi.org/10.1038/s41597-021-01076-6

Chatenoux, B., Richard, J. P., Honeck, E., Mermoz, S., Saah, D., Deslauriers, M., ... & Giuliani, G. (2021). Detecting and monitoring forest degradation using a cloud-based data cube approach. Remote Sensing, 13(22), 4653.

Cao, Q., Li, G., Yao, X., & Ma, Y. (2022). China Data Cube (CDC) for big Earth observation data: Practices and lessons learned. Information, 13(9), 407. https://doi.org/10.3390/info13090407

Cube.js (2021). Cube.js: The open-source framework for building data applications.Cube.js. Retrieved from https://cube.dev/

Building an Earth Observations Data Cube: Lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). (n.d.). CoLab. Retrieved September 27, 2024, from https://colab.ws/articles/10.1080%2F20964471.2017.1398903

Ferreira, K. R., Santos, R., & Queiroz, G. R. (2022). Building Earth observation data cubes with cloud computing: A case study in Brazil using the Open Data Cube platform. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information

Sciences, V-4–2022, 153–159. https://doi.org/10.5194/isprs-annals-V-4-2022-153-2022

Giuliani, G., Chatenoux, B., Honeck, E., & Richard, J. P. (2022). Think global, cube local: an Earth Observation Data Cube's contribution to the Digital Earth vision. International Journal of Digital Earth, 15(1), 1-20.

Giuliani, G., Masó, J., Mazzetti, P., Nativi, S., & Zabala, A. (2019). Paving the way to increased interoperability of Earth observations data cubes. Data, 4(3), 113.

Gomes, V. C. F., Carlos, F. M., Queiroz, G. R., Ferreira, K. R., & Santos, R. (2021). Accessing And Processing Brazilian Earth Observation Data Cubes With The Open Data Cube Platform. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-4–2021, 153–159. https://doi.org/10.5194/isprs-annals-V-4-2021-153-2021

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Guan, H., Liu, H., & Wu, X. (2015). Hybrid parallel computing architectures for geospatial data processing: A case study in remote sensing applications. Computers & Geosciences, 76, 23-34. https://doi.org/10.1016/j.cageo.2014.12.011

Hu, Y., Zhang, Y., & Liu, J. (2018). The increasing volume of geospatial raster data: Challenges and solutions for data management and processing. *Remote Sensing*, 10(4), 611. https://doi.org/10.3390/rs10040611

International Water Management Institute. (2024, May 3). Floods in Kenya underscore urgent need for anticipatory action in disaster situations. International Water Management Institute (IWMI). https://www.iwmi.cgiar.org/news/floods-in-kenya-underscore-urgent-need-for-anticipatory-action-in-disaster-situations/

Kenya Forestry Research Institute. (2018). Strategies for Mango Production in Kenya.

Kenya News Agency. (2023). Satellite data to address food security

Kilimo, R. K. (2014). Land Cover Changes and Landslide Occurrence: A Case of Tirap Division in Elgeyo Marakwet County, Kenya.

Killough, B. (2019). Overview of the open data cube initiative. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 5492-5495). IEEE.

Kang, Y., Zhang, L., & Wang, S. (2019). Computational challenges in geospatial big data analysis: A review and future directions. Big Earth Data, 3(1), 1-20. https://doi.org/10.1080/20964471.2019.1577645

Lenka, S., Parida, V., Sjödin, D. R., & Wincent, J. (2016). Digitalization and its impact on business model innovation: A systematic review of the literature. Journal of Business Research, 69(5), 1868-1875. https://doi.org/10.1016/j.jbusres.2016.01.007

Lewis, B., Killough, B., Purss, M. B., & Siqueira, A. (2019). CEOS Analysis Ready Data (ARD) for Land: An overview on status and future directions. Remote Sensing of Environment, 235, 111485. https://doi.org/10.1016/j.rse.2019.111485

Li, Z., Zhang, L., & Chen, Y. (2016). Geospatial web services for interoperability among heterogeneous datasets: A case study of Earth observation data. ISPRS International Journal of Geo-Information, 5(3), 36. https://doi.org/10.3390/ijgi5030036

Ma, L., Ma, X., Xu, Y., & Zhang, X. (2020). A cloud-native big Earth observation data management platform for spatiotemporal remote sensing analysis. International Journal of Digital Earth, 13(3), 335-353. https://doi.org/10.1080/17538947.2019.1674776

Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski, A., Wojtkowski, C., Krupiński, M., Krätzschmar, E., & Schauer, P. (2020). Forecasting Spatio-Temporal Dynamics on the Land Surface Using Earth Observation Data - A Review. Remote Sensing, 12(21), 3534.

Makueni County Government. (2021). Makueni County spatial planning policy. Makueni County Government. https://www.makueni.go.ke/spatial-planning-policy-2021

Omondi, P. (2022). Scientific and Indigenous Knowledge Understanding of Rainfall Induced Landslides in Murang’a County, Kenya. East African Journal of Environment and Natural Resources, 5(1).

Owers, M., Smith, J., & Jones, P.A.(2022). Advancements in open-source tools for geospatial data processing and analysis: A review of recent developments.Geospatial Data Science Review, 15(2), 123-145.

Rizvi, S. R., Killough, B., Cherry, A., & Gowda, S. (2019). The ceos data cube portal: A guide to open source cloud technology. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 5881-5884). IEEE.

Shao, Y., Wang, J., & Zhang, Y. (2017). Cloud computing for Earth observation data processing: A review of the current status and future prospects. Remote Sensing, 9(8), 827. https://doi.org/10.3390/rs9080827

Soliman, A., & Terstriep, J. (2020). Real-time processing of Earth observation data: Challenges and opportunities in disaster management applications. International Journal of Applied Earth Observation and Geoinformation, 86, 102031. https://doi.org/10.1016/j.jag.2019.102031

Strobl, P., Baumann, P., Lewis, A., Szantoi, Z., Killough, B., Purss, M., ... & Dhu, T. (2023). The six faces of the data cube. In Big Earth Data (pp. 1-26). CRC Press.

VoPham, T., Dwyer, J., & Dwyer,(2018). GeoAI: The role of artificial intelligence in geospatial analysis and its implications for environmental monitoring and management strategies in the future landscape of Earth observation data analysis systems.Environmental Science & Technology, 52(18), 10489-10495.

Wang, S., Hu, F., Li, H., Li, S., & Shi, W. (2023). A Cube-enabled Cloud Geoprocessing Engine for Big Earth Data. International Journal of Digital Earth, 16(1), 1-22.

Xia, H., Huang, Q., Zhu, J., & Guan, Q. (2023). Histogram cube: towards lightweight interactive spatiotemporal aggregation of big earth observation data. International Journal of Digital Earth, 16(1), 2356-2377.

Appendix A: Data Collection Questionnaire to Earth Observation Data Users

Background:

This questionnaire aims to gather information about your organization's use of Earth Observation data and related requirements. Earth Observation data refers to information collected about Earth's physical, chemical, and biological systems via remote sensing technologies. Your responses will help us better understand current usage patterns, challenges, and future needs in the Earth Observation data user community. Please complete all sections as thoroughly as possible.

1. Organization Information

- Name of Organization:
- Department/Division:
- Respondent's Role:

2. Current Earth Observation Data Usage

- How frequently does your organization use Earth Observation data?
  [ ] Daily [ ] Weekly [ ] Monthly [ ] Quarterly [ ] Annually [ ] Never

- What are the primary applications of Earth Observation data in your organization? (Select all that apply)
  [ ] Environmental monitoring
  [ ] Urban planning
  [ ] Agriculture
  [ ] Disaster management
  [ ] Forest management
  [ ] Water resource management
  [ ] Other (please specify): ___________

3. Earth Observation Data Sources

- Which sources of Earth Observation data does your organization currently use? (Select all that apply)

[ ] Landsat
[ ] Sentinel
[ ] MODIS
[ ] Commercial satellite imagery (e.g., Planet, Maxar)
[ ] Aerial photography
[ ] LiDAR
[ ] Other (please specify): ___________

- How satisfied are you with the current Earth Observation data sources?
[ ] Very satisfied [ ] Satisfied [ ] Neutral [ ] Dissatisfied [ ] Very dissatisfied

4. Data Resolution and Quality

- What spatial resolution of Earth Observation data does your organization primarily use?
[ ] Low (>30m) [ ] Medium (10-30m) [ ] High (1-10m) [ ] Very high (<1m)

- What temporal resolution is most crucial for your applications?
[ ] Daily [ ] Weekly [ ] Monthly [ ] Quarterly [ ] Annually

- How important is data quality (e.g., atmospheric correction, cloud-free) for your applications?
[ ] Very important [ ] Important [ ] Moderately important [ ] Slightly important [ ] Not important

5. Data Processing and Analysis

- What software or tools does your organization use for Earth Observation data processing and analysis?

____

- Does your organization have the necessary expertise to process and analyze Earth Observation data effectively?
[ ] Yes [ ] No [ ] Partially

6. Data Accessibility and Sharing

- How does your organization currently access Earth Observation data?
[ ] Direct download from providers

[ ] Through a national/regional data hub
[ ] Commercial data providers
[ ] Other (please specify): ___________

- Does your organization share Earth Observation data or derived products with other organizations?
[ ] Yes [ ] No
If yes, please specify the sharing mechanism: ___________

7. Challenges and Gaps

- What are the main challenges your organization faces in utilizing Earth Observation data? (Select all that apply)
[ ] Data accessibility
[ ] Data quality
[ ] Technical expertise
[ ] Processing capabilities
[ ] Cost
[ ] Other (please specify): ___________

- Are there any specific types of Earth Observation data or products that your organization needs but currently cannot access or use effectively?

____

8. Future Needs and Expectations

- What improvements in Earth Observation data or services would be most beneficial for your organization?

____

- How interested is your organization in using an Earth Observation Data Cube (EODC) system?
[ ] Very interested [ ] Interested [ ] Neutral [ ] Not interested [ ] Not sure

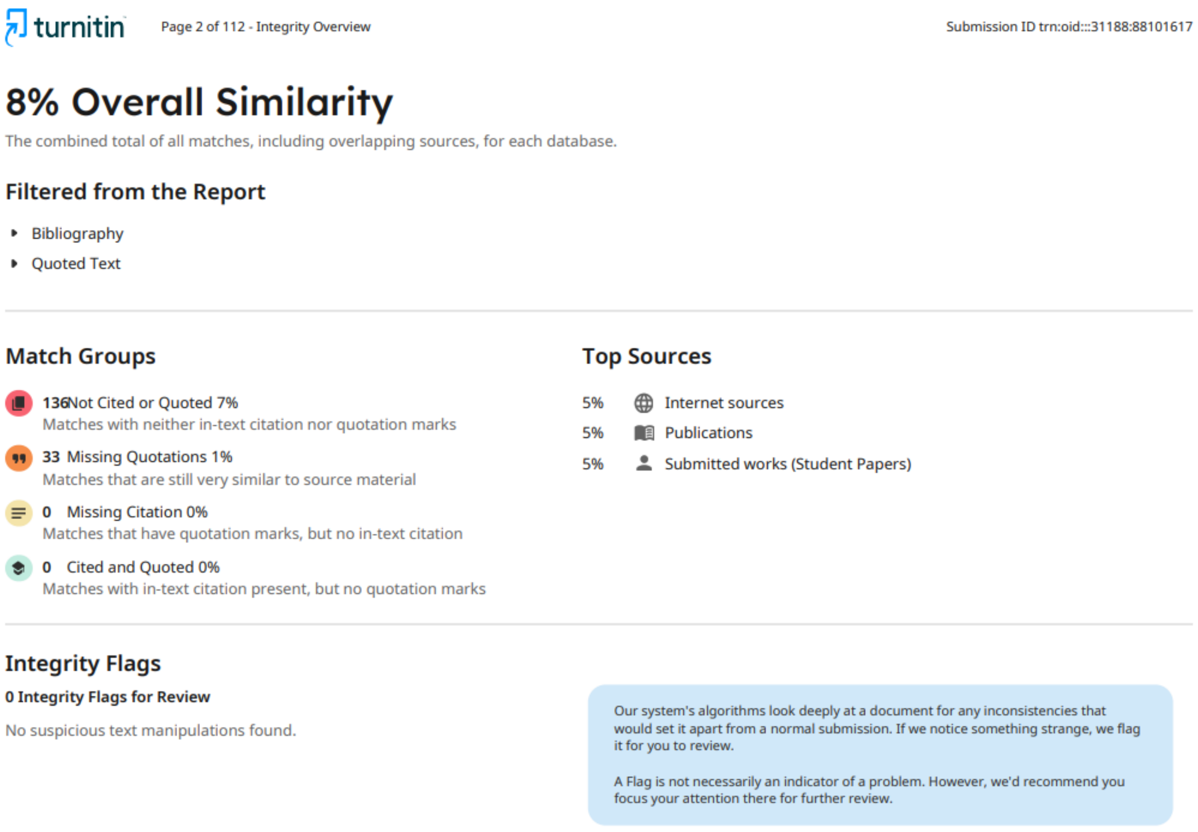
9. Training and Capacity Building

- What type of training or capacity building does your organization need to better utilize Earth Observation data?

10. Additional Comments

- Please provide any additional comments or suggestions regarding Earth Observation data usage and requirements in your organization:

Appendix B: Turnitin Similarity Report



Page 2 of 112 - Integrity Overview

Submission ID trn:oid:::31188:88101617

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

136 Not Cited or Quoted 7%
Matches with neither in-text citation nor quotation marks

33 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5%	Internet sources
5%	Publications
5%	Submitted works (Student Papers)

Integrity Flags

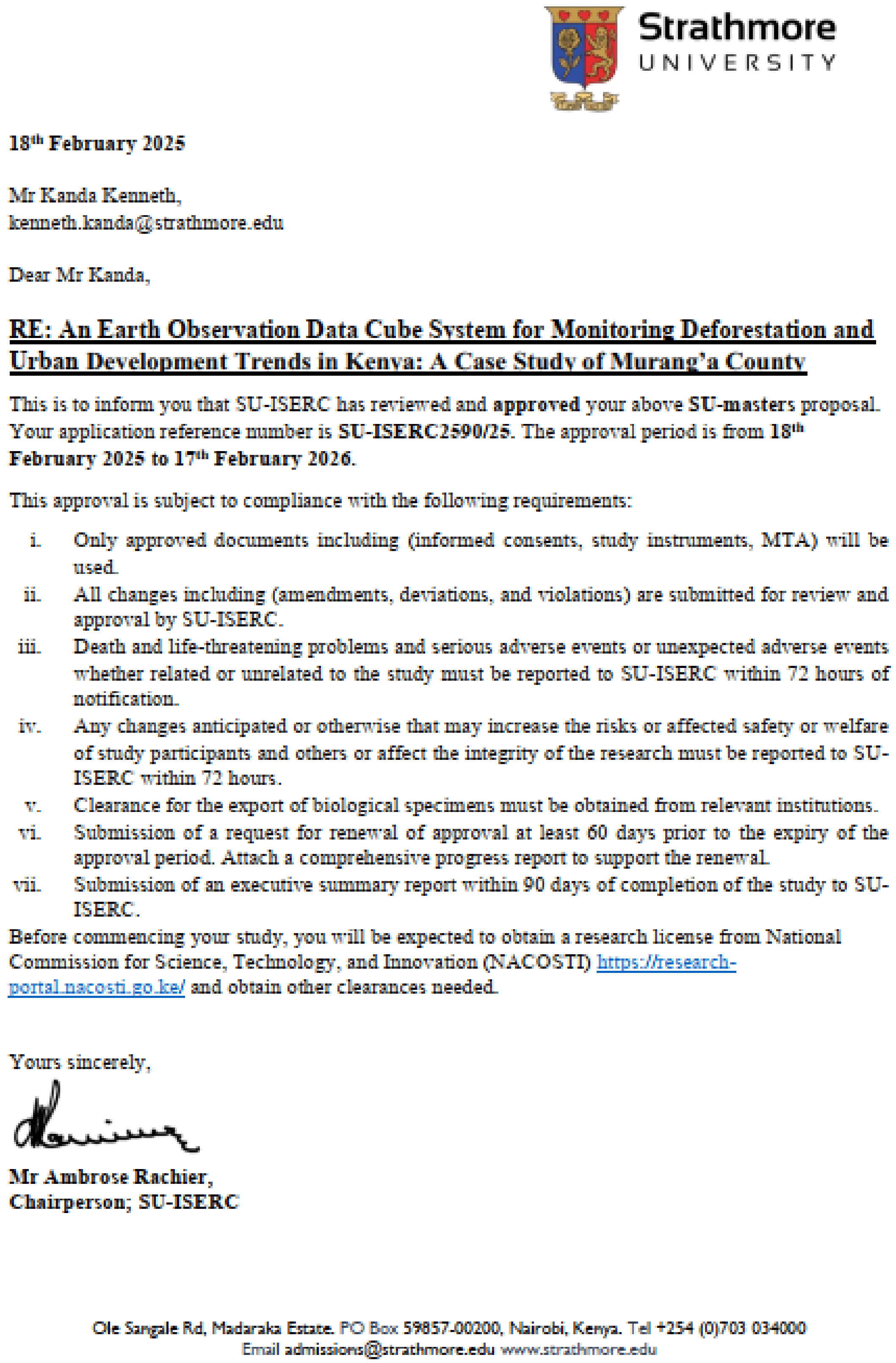
0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Appendix C: Ethics Approval



Strathmore
UNIVERSITY

18th February 2025

Mr Kanda Kenneth,
kenneth.kanda@strathmore.edu

Dear Mr Kanda,

RE: An Earth Observation Data Cube System for Monitoring Deforestation and Urban Development Trends in Kenya: A Case Study of Murang'a County

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2590/25**. The approval period is from **18th February 2025 to 17th February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

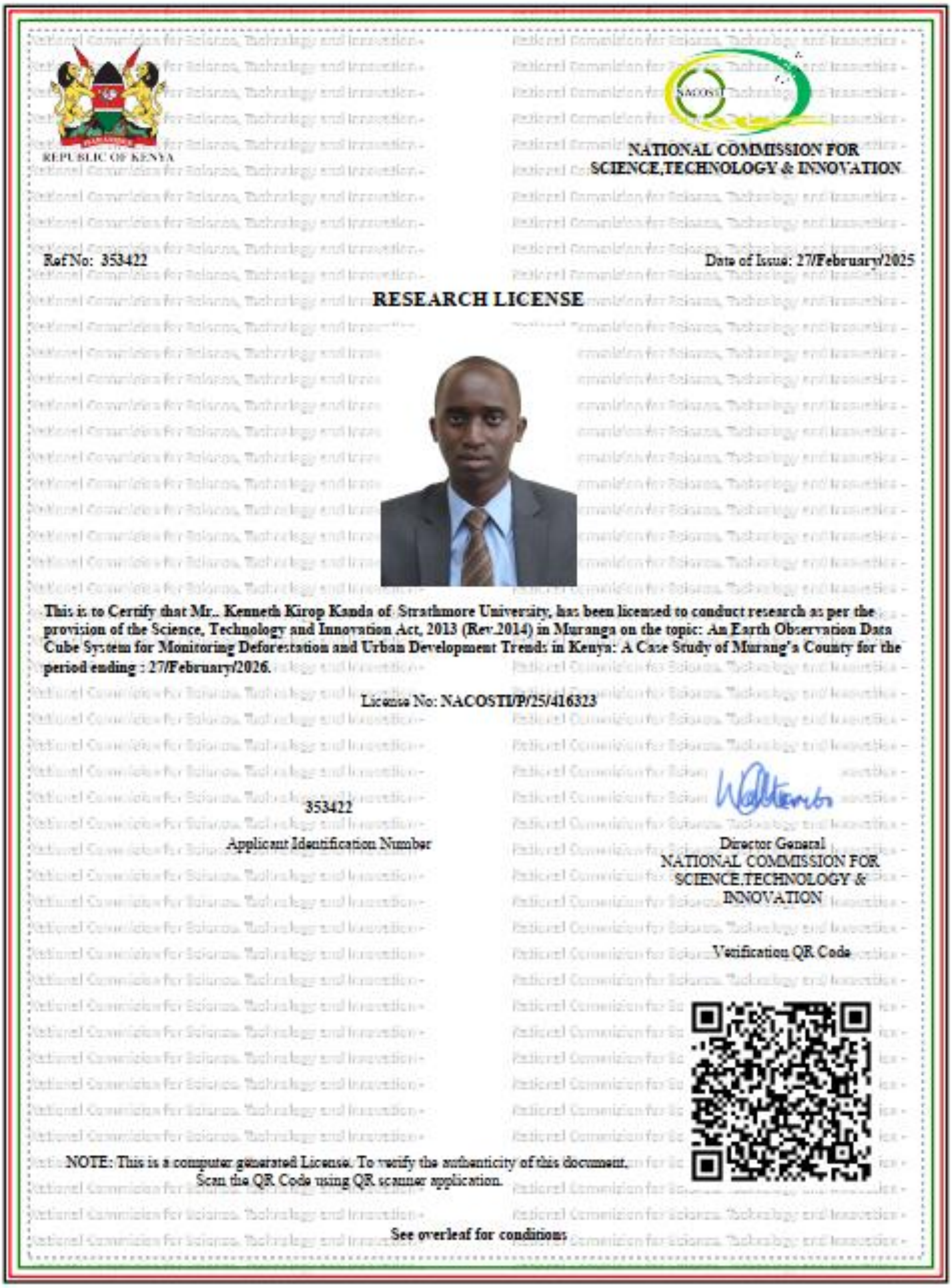
Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix D: NACOSTI Research License



REPUBLIC OF KENYA

NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 353422

Date of Issue: 27/February/2025

RESEARCH LICENSE

This is to Certify that Mr.. Kenneth Kirop Kanda of Strathmore University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Muranga on the topic: An Earth Observation Data Cube System for Monitoring Deforestation and Urban Development Trends in Kenya: A Case Study of Murang'a County for the period ending : 27/February/2026.

License No: NACOSTI/P/25/416323

353422

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY &
INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document,
Scan the QR Code using QR scanner application.

See overleaf for conditions