Research Article

pISSN 1225-6161 • eISSN 2287-9307
Korean Journal of Remote Sensing 2025, 41(5), 859-868
https://doi.org/10.7780/kjrs.2025.41.5.12

Korean Journal of Remote Sensing
KJRS

Conceptual Design of an XCube-Based Data Cube for Geospatial Applications with Remotely Sensed Imagery

Seungjae Joo[1], Kwangseob Kim[2], Kiwon Lee[3]*

[1]Master Student, Department of Convergence Security, Hansung University, Seoul, Republic of Korea
[2]Assistant Professor, Department of Computer Software, Kyungmin University, Uijeongbu, Republic of Korea
[3]Professor, Department of Convergence Security, Hansung University, Seoul, Republic of Korea

Abstract: This study presents the conceptual design of a multidimensional data cube for remote sensing applications, which was developed using the open-source XCube framework. Unlike traditional file-based or relational database systems for data storage, and data hubs or catalogs for data services, the proposed structure incorporates spatial, temporal, spectral, and sensor dimensions, facilitating efficient large-scale spatiotemporal analysis. The framework supports the scalable and interoperable management of very high-resolution Earth observation (EO) datasets, such as Compact Advanced Satellite 500-1 (CAS500-1) imagery, using Zarr-based cloud storage. This study demonstrates interoperability with platforms, including Google Earth Engine (GEE), and focuses on region-specific data sovereignty. Two pilot cases - assessing risk levels in densely built, aged residential areas and analyzing priority areas for road maintenance in forest degradation zones - exemplify the practical utility of this approach in environmental analysis and policy support. The results suggest the potential of an XCube-based data cube to provide a flexible, scalable infrastructure for geospatial analytics.

Keywords: Cloud computing, Data cube, Google Earth Engine, XCube, Zarr

Received: September 24, 2025
Revised: October 21, 2025
Accepted: October 29, 2025
Published: October 31, 2025

Corresponding author:
Kiwon Lee
E-mail: kilee@hansung.ac.kr

1. Introduction

In recent years, there has been an increasing focus on multidimensional spatiotemporal analysis frameworks in remote sensing-based geospatial analysis. These frameworks can consider spatial distribution and temporal dynamics simultaneously. Traditional geospatial approaches have mainly relied on spatial data at a single time point or on simple time-series analyses. This provides limited interpretability for advanced applications such as change detection, multi-sensor data fusion, and climate change prediction. In modern analytical environments, the integration and harmonized analysis of multi-sensor satellite imagery, such as Sentinel-2, Landsat series, and Compact Advanced Satellite 500-1 (CAS500-1), are required; conventional frameworks face substantial challenges due to inconsistent spatial/temporal resolutions, spectral band heterogeneity, and differences in data storage structures.

Augustin et al. (2019) designed and implemented the Earth System Data Cube (ESDC) infrastructure for a data-cube theme across multiple geospatial datasets. The ESDC recognizes multi-resolution data from diverse sensors, including the Sentinel, Moderate Resolution Imaging Spectroradiometer (MODIS), and Landsat series, and normalizes it into a grid structure. Appel and Pebesma (2019) introduced the gdalcubes library, an R package for constructing on-demand data cubes. They presented a case study using Sentinel-2 time-series data to demonstrate its functionality. Mahecha et al. (2020) built on this foundation and emphasized the necessity of data cubes in Earth system science. They noted that, despite the massive volume of spatiotemporal data collected from numerous satellites and sensors, its heterogeneity, complexity, and lack of standardization significantly hinder efficient analysis and utilization. Giuliani et al. (2020) further highlighted that, by storing high-dimensional, multivariate datasets in a normalized, grid-based format, data cubes provide a straightforward pathway to generating standardized structures well-suited for artificial intelligence (AI)-driven analysis. Similarly, Kasprzyk and Devillet (2021) described an advanced analytical technique for efficiently exploring and analyzing large-scale spatiotemporal data that is strongly relevant to data cube methodologies. Building on these developments, Cao et al. (2022) and Gao et al. (2022) proposed a data cube model that integrates remote sensing data from multiple sensors with varying resolutions and temporal frequencies into a unified structure. This enables large-scale geospatial analysis. Montero et al. (2024) noted that data cubes are flexible structures that can be integrated with image collections and information-preserving cube systems, being interoperable with platforms such as Google Earth Engine (GEE) and Open Data Cube (ODC). Yue et al. (2024) addressed the growing need for AI-ready geospatial data and noted that Earth observation (EO) data requires extensive preprocessing and structuring before it can be used to train AI models, which introduced an open-source toolkit that automatically transforms conventional geospatial data cubes into AI-ready data cubes optimized for machine learning and AI-based analysis.

Building on the Australian Data Cube (ADC) to ensure the long-term management and utilization of satellite-based EO data (Lewis et al., 2017; Dhu et al., 2019), this initiative expanded internationally through the Committee on Earth Observation Satellites (CEOS) under the Data Cube Initiative (Killough, 2018) and evolved into the open-source ODC platform. The ODC enhances large-scale spatiotemporal analysis by integrating multi-sensor and multi-temporal datasets while leveraging cloud infrastructures to improve the efficiency of processing large volumes of data. It also enables countries to develop data cubes tailored to their specific environmental contexts. The ODC is currently in use across Europe, Asia, Africa, and South America (Chatenoux et al., 2021; Cao et al., 2022; Sudmanns et al., 2023; Kim and Lee, 2023; Gomes et al., 2024). According to the Post-2025 Group on Earth Observations (GEO) Work Program Summary Document (Group on Earth Observations, 2025), the ODC is expected to continue as part of the Digital Earth Approaches for GEO—Data Cube and Open Platforms (DECOP), positioned within the Enabling Mechanisms framework to promote connection and integration across the GEO community.

The ODC provides a robust foundation for managing multi-sensor and multi-temporal EO data. However, XCube (https://xcube.readthedocs.io/en/latest/) has been reported to complement this architecture. It enhances scalability for large-scale data processing, strengthens spatio-temporal analytical capabilities, and improves interoperability with cloud-native geospatial infrastructures (Baumann et al., 2019). Through this integration, XCube enables the efficient handling of massive satellite datasets and provides a basis for incorporating multidimensional data streams, including atmospheric, spectral, and climate-related variables (Web-based European Knowledge and Earth Observation, 2023). The Euro Data Cube (EDC) is a cloud-based EO data analytics platform jointly developed and operated by the European Space Agency (ESA) and several partner organizations. It adopts the XCube framework (https://eo4society.esa.int/resources/euro-data-cube/). While the EDC builds on a data cube concept similar to the CEOS ODC, it is distinguished by its close integration with the Copernicus program.

The goal of this research is to develop a multidimensional data cube concept for satellite remote sensing imagery. This concept would enable the integrated management and further analysis of large-scale spatiotemporal data within an open-source, cloud-based computing environment such as OpenStack. The study proposes a conceptual design for a data cube based on XCube that enables the efficient management and analysis of high-resolution EO data, such as CAS500-1 imagery, by providing cloud-optimized Zarr-based storage. The study addresses fundamental technical aspects by comparing databases and data cubes and examining their differences with other data service mechanisms, such as data hubs and data catalogs. Within the conceptual design process, the study proposes a linkage model with GEE. Ideally, implementing such a conceptual model would require diverse types of satellite imagery and other grid-based datasets collected over multiple periods. However, this research presents two simplified use cases that utilize limited data. This study does not aim to invent a new data cube technology; rather, it emphasizes the strategic adaptation and localization of a powerful open-source framework, XCube, to address national data sovereignty priorities by leveraging freely distributed domestic satellite assets, such as CAS500-1. The novelty of this work lies in its socio-technical approach, which deploys a modern, cloud-native stack that bridges global open-source data cube initiatives with region-specific geospatial infrastructure for local government. In this context, the proposed framework presents use cases as a feasibility study for implementing the "Think global, cube local" paradigm, serving as a practical reference model for developing tailored data cube architectures.

2. Materials and Methods

2.1. Basic Concepts of Data Cube

Fig. 1 provides a clear conceptual representation of the multi-dimensional data cube architecture, which is widely used in EO data manipulation and geospatial analysis systems to efficiently manage, store, and analyze large-scale spatiotemporal datasets. Each data cube, labeled Data Cube A and Data Cube B, is a four-dimensional array comprising spatial (x, y) and temporal (time) dimensions, as well as a spectral or variable (band) dimension.

This structure enables simultaneous analysis across space, time, and observed variables for applications such as vegetation monitoring, climate analysis, and land-use/land-cover (LULC) classification. The data cubes represent different variables or thematic layers, such as normalized difference vegetation index (NDVI), temperature, and precipitation, and are integrated within a data array that serves as the central analytical structure. Separating variables into distinct cubes allows for modular analysis while supporting coordinated processing across multiple data layers. Datasets are the organizational level that aggregates multiple data cubes into a cohesive analytical collection. This hierarchical structure is common in EO platforms such as ODC, XCube, and GEE, where multiple variables are bundled together to support multidimensional and multivariate analysis.

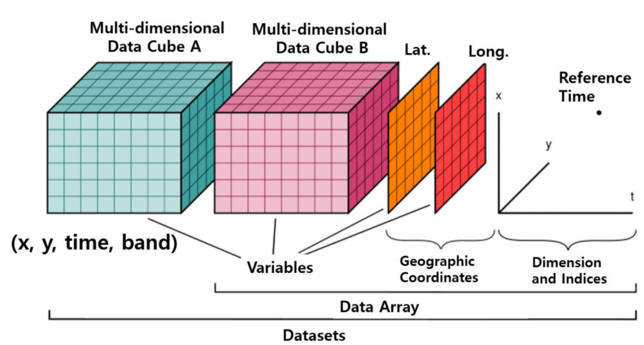

Fig. 1. Concept of data cube, edited from Hoyer and Hamman (2017).

To ensure proper geospatial referencing, the system uses latitude and longitude grids, shown as two-dimensional slices in the diagram. These geographic coordinates establish the spatial framework for all data points. The underlying coordinate system enables analysts to relate pixel values to real-world locations. The x, y, and t axes indicate the dimension indices that define the data cube's structure. These indices are used for indexing, slicing, and querying specific spatial or temporal subsets. A reference time is also provided, serving as a temporal anchor for synchronizing multi-temporal datasets or initiating time-series analysis.

Recent developments have focused on cloud-optimized data formats such as Zarr and Cloud-Optimized GeoTIFF (COG), as well as on data cube-based platforms such as XCube, ODC, and GEE. Zarr is an open standard for storing large-scale, multidimensional array data that allows data to be saved in array chunks. A COG is a GeoTIFF file that is internally arranged to support efficient data access and processing in cloud-based workflows.

Spatial information databases (DBs) and geo-based data cubes serve distinct yet complementary roles in geospatial data management and analysis (Table 1). Spatial information databases are primarily designed to store, query, and analyze spatial objects, focusing on vector or raster data within a relational database framework, such as PostGIS. They are typically used in land registration, urban planning, and facility management. They rely on SQL-based spatial operations and are often constrained to discrete time attributes. In contrast, spatial data cubes are optimized for time-series analysis, large-scale data processing, and AI-driven applications. They adopt a multidimensional array structure that enables analysis across spatial (x, y), temporal (time), and spectral (band) dimensions. This structure supports complex analyses, such as satellite image interpretation, climate modeling, and NDVI-based vegetation change detection. Importantly, data cubes treat time as a core analytical dimension, enabling seamless integration of time-series statistics and machine learning methods. Technologically, spatial information DBs commonly use platforms

Table 1. Comparison of the spatial information database and the geo-based data cube

Item	Spatial Information Database	Geo Data Cube
PurposeC	Mainly for storing, searching, and analyzing spatial objects	Supports time-series analysis, big data analysis, and AI learning
Structure	Vector/Raster-based relational DB	Multidimensional array structure (x, y, time, bands)
Data Type	Vector data such as roads, buildings, and administrative areas	Time-series raster data such as satellite and climate images
Application Field	Land registration, urban planning, and facility management	Satellite image analysis, climate modeling, NDVI change detection
Time Dimension	Time stored only as unit time or attribute	Includes time dimension as a basic element
Analysis Method	SQL-based queries and spatial operations	Supports time-series statistics and machine learning-based analysis

Table 2. Comparison of Data Hub, Data Catalog, and Data Cube

Category	Data Hub	Data Catalog	Data Cube
Purpose	Stores and shares data from various sources	Manages data asset lists and metadata	Structures data for multidimensional analysis
Focus	Integration and sharing	Exploration and search	Systematic time management, analysis, and visualization
Function	Provides data collection, transformation, integration, and API	Provides metadata such as location, description, owner, and quality	OLAP-based multidimensional analysis
Form	Centralized data storage	Metadata management system	Structured analytical database

such as PostGIS, Oracle Spatial, and ArcGIS Geodatabase. In contrast, data cubes use tools such as ODC, XCube, and GEE to handle large spatiotemporal datasets more efficiently.

In the realm of data infrastructure, Data Hubs, Data Catalogs, and Data Cubes represent three complementary systems with distinct purposes, functional scopes, and user bases (Table 2). Data Hubs are primarily designed to store and share data collected from diverse sources. Their primary focus is on integration and sharing, enabling users to access a centralized data repository via standardized interfaces such as application programming interfaces (APIs). They typically support processes like data collection, transformation, and integration. These platforms are widely used by data engineers and platform operators, and often underpin national data infrastructures, such as public data portals and integrated APIs.

Data Catalogs, by contrast, serve as metadata management systems. They facilitate the exploration and discovery of datasets by maintaining comprehensive metadata, including location, description, ownership, and quality. Data Catalogs are essential for organizing and indexing data assets, making them accessible to data managers and analysts. Data cubes go a step further by structuring data for multidimensional analysis, particularly across spatial and temporal dimensions. Their key strength lies in systematic time-series management, analysis, and visualization, often supported by Online Analytical Processing (OLAP) capabilities. These platforms are optimized for advanced analytics, such as regional population dynamics and crime trend analysis, and are primarily used by policy makers, analysts, and public administrators. The data cube model facilitates decision-making by integrating large-scale data into a coherent, analytical framework.

Recent advancements in EO-based geospatial analysis have led to the development of specialized platforms, including XCube, GEE, and ODC. Table 3 summarizes the key characteristics of data cube-based technologies, including XCube and ODC. These systems support the growing demand for scalable, time-aware analysis of massive spatiotemporal datasets. Although these platforms serve similar domains, they differ significantly in design philosophy, accessibility, and analytical capabilities.

Table 3. Comparison of XCube, GEE, and ODC

Item	XCube	Google Earth Engine (GEE)	Open Data Cube (ODC)
Data Access Method	Local/Cloud-based Zarr Cube	Cloud API	Local DB or Cloud Storage
Programming Language	Python	JavaScript, Python	Python
UI/Visualization	XCube Viewer, Jupyter	GEE Code Editor	Jupyter, QGIS
Data Provision	External manual import	Automatic via Landsat, Sentinel, MODIS, etc.	Users build manually
Installation/Use	Requires Python installation	Web-based, No installation needed	Complex installation and setup
Target Users	Researchers, developers, and general users	General users to experts	Mainly public institutions, researchers, and general users
Scalability	Server, API extensible	Limited	Highly scalable
Main Features	Zarr cube analysis, visualization, API	Large-scale image processing, Machine learning, statistics	National-scale analysis, LULC classification
Strengths	Multidimensional analysis, Xarray compatibility	Fast, efficient, massive datasets	Comprehensive large-area analysis
Weaknesses	Requires manual setup	Google dependency, customization limited	Complex installation, high learning curve

https://doi.org/10.7780/kjrs.2025.41.5.12

The rasdaman system (http://www.rasdaman.org/), recognized as a pioneering Array Database Management System (Array DBMS), laid the foundational concepts for managing and querying large-scale multidimensional raster data through array-oriented operations, and XCube extends this paradigm.

XCube, developed by the German Aerospace Center (DLR), is a Python-based platform that provides local or cloud-based access to Zarr cube datasets. It is ideal for advanced time series analysis, particularly when used with Xarray (Hoyer and Hamman, 2017), and supports interactive visualization with XCube Viewer or Jupyter Notebooks. However, it requires manual installation and setup, making it better suited to researchers and developers with Python expertise. The Pangeo ecosystem (https://pangeo.io/) provides a foundation for scalable geoscientific data analysis in cloud environments. XCube builds upon this foundation, extending Pangeo's capabilities by enabling efficient generation and management of spatiotemporal data cubes.

GEE is a fully cloud-based geospatial analysis platform that supports JavaScript and Python. It provides automatic access to a wide range of satellite data, such as Sentinel and MODIS, enabling fast, large-scale image processing and statistical analysis. GEE offers a web-based development environment with a code editor, making it user-friendly for beginners and experts alike. Despite its strengths in speed and accessibility, GEE's closed ecosystem and limited customization options are potential drawbacks for advanced users (Amani et al., 2020). Instead, GEE is currently one of the most widely used platforms for practical satellite image processing and analysis, and it shares similarities with data cube approaches while exhibiting differences (Gomes et al., 2020).

The ODC community collaborates with organizations such as Geoscience Australia to maintain ODC. This open-source platform supports robust time-series processing and the structured storage of large satellite datasets via local databases or cloud infrastructure. ODC is implemented in Python and is often integrated with tools such as Jupyter and QGIS. ODC is particularly well-suited for national-scale analyses and LULC classifications. However, it is relatively complex to install, and the system has a steep learning curve, making it more appropriate for institutional users.

In terms of scalability and analytical capability, both XCube and ODC support multidimensional data models and extensibility via APIs. GEE, on the other hand, offers high performance within a managed environment. The choice among these platforms should be guided by user expertise, infrastructure capacity, and analytical goals.

2.2. Conceptual Design for Applications

This study established a conceptual framework for systematically constructing a multidimensional remote sensing data cube based on the characteristics of geospatial information. The framework consists of three components: defining the structural elements of the data cube, designing the processing flow architecture, and providing an implementation example using the open-source XCube framework.

Fig. 2 shows the workflow for preparing data for a geospatial data cube. In practice, multiple heterogeneous datasets, such as remotely sensed imagery, geospatial information, and open socioeconomic data, must first be integrated into a common analytical framework before being converted into a unified data cube structure.

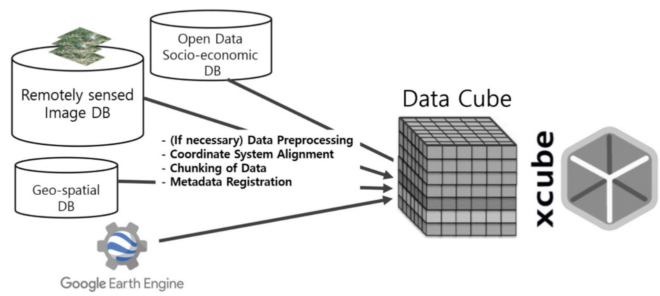

Fig. 2. Work flow and research scheme.

The processing architecture is organized into four sequential stages: data acquisition, preprocessing, cube transformation, and analysis. Sources of EO data are obtained through open APIs or cloud-based platforms, such as Copernicus and Amazon Web Services (AWS) Open Data. During preprocessing, operations such as cloud masking, atmospheric correction, geometric alignment, band selection, and index derivation are performed. Then, the processed data are transformed into the Zarr format, which efficiently represents multidimensional arrays and supports the simultaneous handling of spatial, temporal, spectral, and sensor axes via spatiotemporal indexing and grid alignment. The analysis stage supports various applications, such as computing NDVI time series, detecting spatiotemporal anomalies, providing advanced visualization, and training machine learning models. These functions demonstrate the proposed framework's potential to serve as a practical, scalable infrastructure for environmental monitoring and geospatial analytics.

Several essential preprocessing steps are required to achieve this integration. First, the input datasets must be harmonized. Where necessary, radiometric or spatial preprocessing is performed to ensure consistent spatial resolution and format. Next, all data sources must be reprojected to an identical coordinate reference system to guarantee spatial alignment across different data layers. Once the datasets share a common spatial frame, the data are divided into smaller chunks for efficient storage and processing in parallel using modern array-based architectures such as Zarr and Xarray. Finally, detailed metadata, including spatial/temporal extent, variable descriptors, and sensor information, is registered to allow proper indexing and querying of the resulting data cube.

After completing these preparatory steps, the processed datasets can be ingested into a multidimensional data cube. This cube supports spatiotemporal analysis and visualization on platforms such as XCube or GEE. This workflow ensures that diverse geospatial information is ready for analysis and can be queried systematically across space, time, and variables within a scalable data infrastructure.

Fig. 3 shows the overall workflow for creating and delivering an XCube-based data cube from satellite imagery. In this workflow, the CAS500-1 imagery and other processed datasets from the GEE and NumPy/Xarray environments are first converted into multidimensional arrays ready for analysis. Xarray is particularly appropriate here because it allows direct manipulation of labeled, multifaceted data and supports the Zarr format, which is necessary for scalable data storage.

After converting the input data into Zarr-based chunks, the data are stored in a cloud environment (OpenStack in this case) and ingested into the XCube framework. Within the framework, the XCube Generator creates the data cube from the Zarr inputs, and the XCube Server exposes the created cube via standardized APIs. The XCube-viewer allows users to explore and visualize the spatiotemporal data interactively. Additionally, the diagram correctly illustrates the optional use of GeoJSON features, which can be overlaid on the data cube to define regions of interest or support vector-based spatial queries. Overall, the diagram is conceptually accurate and well-aligned with established best practices for constructing scalable geospatial data cubes. It demonstrates a clearly structured process for transforming raw satellite imagery into service-based data delivery using the XCube ecosystem.

The XCube–Earth Engine (XEE) framework is an add-on that connects GEE with the XCube data cube system. With XEE, users can pull GEE datasets directly into an XCube instance, combining the strengths of both systems. XEE enables users to access and ingest GEE datasets directly into an XCube instance, combining the strengths of both platforms. This integration enables the construction of scalable data cubes with standardized Zarr formats and metadata, while ensuring interoperability across cloud services. Practically speaking, XEE enables researchers to leverage GEE's extensive Data Catalog and processing capabilities within the structured, analysis-ready XCube framework, supporting advanced applications in LULC monitoring, climate analysis, and environmental decision-making.

3. Results

The data cube is designed as a multidimensional structure that incorporates spatial, temporal, and spectral characteristics simultaneously. Spatial resolution is typically set to 10 m, while temporal resolution can be defined flexibly according to user needs. For example, it can be set to monthly or 16-day intervals. The spectral dimensions encompass the original sensor bands and derived indices, such as NDVI, SWIR, NIR, and VV/VH polarization data. This enables comparative time-series analysis within the cube.

Consideration was also given to integrating various EO satellite datasets. CAS500-1 provides high-resolution optical imagery for domestic uses; Sentinel-2 offers medium-resolution optical imagery; and the Landsat series contributes long-term time series archives. Despite their differences in spatial resolution,

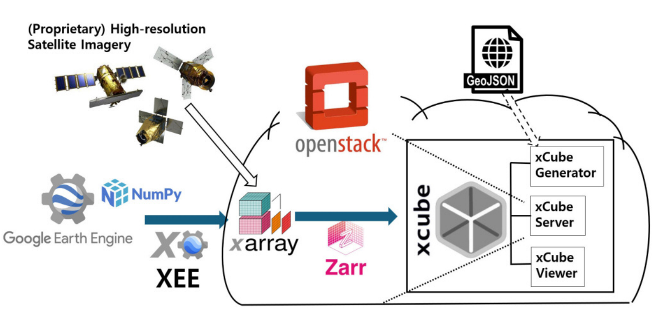

Fig. 3. Conceptual workflow for XCube-based data cube building.

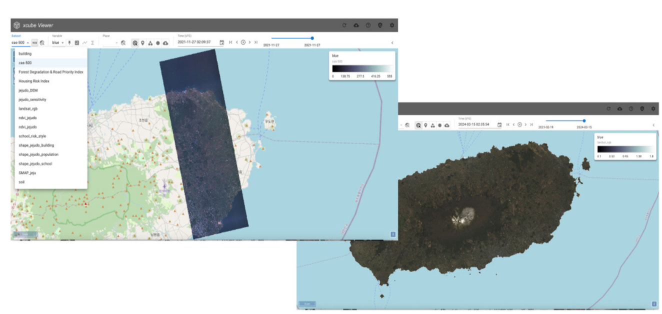

Fig. 4. Jeju data cube as a pilot system, showing CAS500-1 and clipped Landsat imagery on XCube-viewer.

https://doi.org/10.7780/kjrs.2025.41.5.12

Table 4. Datasets for use case applications

No	Data (Date)	Source
1	Land Satellite (CAS500-1) (Oct. and Nov. 2021)	National Geographic Information Institute, National Land Information Platform
2	LANDSAT (Jan. 2022)	Google Earth Engine
3	Jeju Island 100 m grid total population data (April and Oct. 2020)	National Geographic Information Institute, National Land Information Platform
4	NDVI values (Processed from Landsat data) (Jan. 2022)	Google Earth Engine
5	DEM (Digital Elevation Model) (Sept. 2024)	National Geographic Information Institute, National Land Information Platform
6	Geological Data (Nov. 2020)	Public Data Portal (https://www.data.go.kr/data/15073423/fileData.do)
7	Jeju Island 100 m grid building density data (Nov. 2020)	National Geographic Information Institute, National Land Information Platform
8	Jeju Island 100 m grid-aged housing data (Jan. 2020)	Public Data Portal (https://www.data.go.kr/data/15073424/fileData.do)
10	NASA SMAP (Soil Moisture Active Passive) (Feb. 2023)	Google Earth Engine
11	Jeju Island road data (July 2025)	V-world (https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?dsId=30055)

temporal coverage, and spectral configuration, these datasets can be normalized within the cube structure, enabling harmonized, integrated analysis. Fig. 4 shows the initial interface of the Jeju data cube based on XCube, which was provisionally established as a pilot system, and Table 4 lists the datasets registered in this cube.

In this study, the presented use case should be regarded as a proof-of-concept (POC) experiment demonstrating the feasibility of the proposed conceptual framework. The utilization of input datasets from different periods was both inevitable and deliberate as part of the research strategy. In practice, acquiring multi-source datasets with perfectly matched temporal coverage and spatial resolution requires substantial cost and time. Such constraints are particularly evident during the initial stages of conceptualizing national or region-specific data cubes. Accordingly, this study prioritized the use of available datasets to illustrate how heterogeneous spatiotemporal information can be harmoniously managed and analyzed within an integrated data cube structure.

Fig. 5 illustrates the overall analytical workflow with two pilot cases. For this study, a linear weighted combination approach was used to construct a multi-factor index (Eq. 1). Although more sophisticated numerical modeling techniques or advanced algorithms could be applied, this model was developed as an initial framework based on a limited set of available data in the conceptual design stage; therefore, its applicability is limited. The input variables presented in Table 4 were limited by the difficulty of obtaining data within the valid temporal ranges required for rigorous analysis or validation. Accordingly, the purpose of this pilot experiment is to demonstrate a feasible application model, rather than to provide fully substantiated or generalizable results, as shown in Eq. (1).

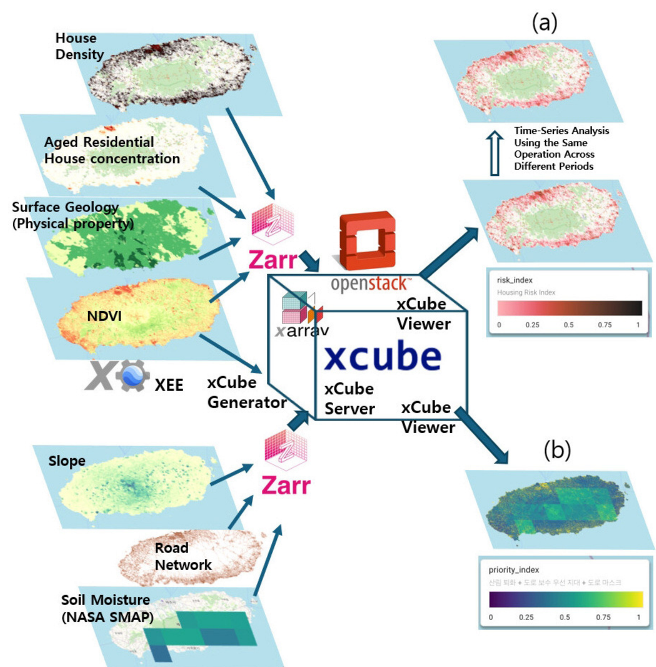

Fig. 5. Process pipeline of use cases. (a) Risk level of densely built aged residential areas. (b) Analysis of priority areas for road maintenance in forest degradation zones.

$$\text{Expected value as index} = \alpha \cdot A + \beta \cdot B + \gamma \cdot C + \delta \cdot D + \cdots \quad (1)$$

A, B, C, and D mean variables in the data cube, and α, β, γ, and δ are weights. That is, the input variables A, B, C, and D may correspond to the datasets presented in Table 4, and, depending on the application purpose, each variable can be assigned a weight ranging from 0 to 1. In particular, the data cube primarily provides the technical foundation for combining and integrating heterogeneous datasets rather than directly generating individual data for a specific application. Therefore, it is common practice to use verified datasets from official data providers rather than creating new input data manually. Consequently, the data cube does not directly handle error detection or validation of individual input variables. The datasets presented in Table 4 are derived from verified data sources. Compared to other input datasets, the Digital Elevation Model (DEM) and road data were obtained from recent sources in 2024 and 2025. Although these two datasets may continue to undergo detailed local changes, it was assumed that, at the regional scale of Jeju Island, there have been no significant variations since the acquisition of the socio-demographic data in 2020 and the satellite imagery in 2022. Therefore, these datasets were applied to this POC study under that assumption.

Fig. 5(a) is a test case to evaluate housing-related risks in densely populated, aging residential areas. In this framework, multiple spatial variables, including house density, aged residential house concentration, surface geology for physical properties, and NDVI, are first integrated into a multidimensional data cube using the XCube-generator. Using the Zarr format and Xarray ensures the data is stored efficiently as analysis-ready arrays, and the OpenStack-based storage environment provides scalable infrastructure. Once constructed, the data cube is served by the XCube Server, enabling on-demand access to all spatial and temporal layers. The generated cube can then be visualized interactively in the XCube Viewer. There, a composite risk index such as "Risk Level of Densely Built Aged Residential Areas" is calculated by applying a common operation across the different variable layers. Furthermore, the proposed approach supports time-series analysis by applying the same operation to different time periods, enabling tracking of spatiotemporal changes in risk levels. This visualization is conceptually valid and methodologically appropriate because it clearly and logically represents how heterogeneous geospatial datasets can be integrated, processed, and analyzed within a data cube environment to derive meaningful, policy-supporting indicators.

Fig. 5(b) is an example to identify priority areas for road maintenance in forest degradation zones. In this example, the XCube generator ingests multiple environmental and infrastructural variables, including slope, road network, soil moisture, and NDVI, and stores them as Zarr-based multidimensional arrays in an Xcube-based data cube. This allows for the handling of heterogeneous datasets within a unified spatial and temporal framework supported by OpenStack's scalable cloud infrastructure. The generated cube is then served through the XCube server, enabling interactive exploration and analysis within the XCube viewer. A priority index is calculated by combining the individual variables in a spatially consistent manner. This allows critical areas requiring road maintenance in forest degradation zones to be clearly identified. The resulting visualization highlights spatial patterns of road maintenance priorities, demonstrating the proposed approach's capability to support evidence-based decision-making in forest management.

The datasets presented in Fig. 5 are all managed using a Zarr-based data storage system, and a portion of them is accessed via integration with GEE. This integration indicates that the proposed framework is technically capable of providing an environment that ensures robust interoperability with cloud-native infrastructures and the GEE ecosystem by employing an optimized Zarr-based data storage architecture that supports scalable, chunked, and parallelized access to multidimensional satellite datasets. Furthermore, as demonstrated in the POC example, if time-series datasets are continuously accumulated and systematically managed within this framework, it will enable the system to operate effectively within the Korean geospatial data ecosystem and present strong potential for integration with satellite imagery–based administrative decision-making processes at the local government level.

Meanwhile, as new datasets continue to be produced for the Jeju region, the Jeju data cube developed in this study will also accumulate additional time-series data over time. Such regional data cubes can be directly applied or extended to other regional or thematic data cube systems. As data accumulation continues, the development of satellite image time-series analysis models based on the data cube framework will advance.

4. Conclusions

The primary objective in this study is to propose a conceptual framework for an XCube-based data cube for regional applications. The proposed framework builds on the shared conceptual foundation of global initiatives or country-wide scale data cubes, such as the ODC and EDC, but emphasizes several distinctive contributions. First, it focuses on domestic, high-resolution datasets, such as CAS500-1. This addresses region-specific data sovereignty, which is less emphasized in the ODC and EDC. Second, the framework demonstrates interoperability with cloud-native infrastructures and GEE while utilizing Zarr-based storage. Third, the proposed framework demonstrates strong potential for long-term scalability and sustainability. As time-series datasets continue to be accumulated and systematically managed, it is expected to function effectively within the future Korean geospatial ecosystem, established by government agencies or the private sector, and support satellite imagery–based administrative decision-making at the local government level.

These characteristics highlight its novelty as a national-level adaptation that builds on, rather than replicates, prior international efforts. Thus, this study is to emphasize the necessity of establishing a modern cloud-native stack that strengthens data sovereignty by strategically adapting and localizing the powerful open-source framework XCube while utilizing domestic satellite assets such as CAS500-1.

Although the presented use cases successfully demonstrate the feasibility and flexibility of the proposed framework, they rely on datasets collected from different temporal periods. This limitation was an intentional and pragmatic choice to verify the conceptual applicability of the framework at the preliminary stage. Consequently, the derived indices, such as the risk level and maintenance priority, can be regarded as illustrative outputs that exemplify the workflow's analytical process rather than analytically validated results. Future research will focus on acquiring and processing temporally harmonized, multi-source datasets to rigorously validate the proposed model. This includes applying machine learning–based evaluation methods and quantitative accuracy assessments to establish reproducible, scalable benchmarks for operational deployment.

Building on this foundation, future research will develop more rigorously harmonized datasets, enabling the application of comprehensive validation metrics and machine-learning–based evaluation methods. In this sense, the present work is significant as an initial, essential contribution that lays the groundwork for advancing toward more robust, scalable, and operational implementations of the proposed data cube framework.

Author Contributions

Conceptualization: Lee K, Kim K; Data curation, Formal analysis: Lee K, Joo S; Funding acquisition: Lee K , Investigation, Methodology: Lee K, Kim K, Joo S; Project administration: Lee K; Supervision: Lee K; Validation: All authors; Writing–original draft: Lee K, Joo S; Writing–review & editing: All authors.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Funding

This research was financially supported by Hansung University.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

None.

Supplementary Materials

None.

References

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., et al., 2020. Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 5326–5350. https://doi.org/10.1109/JSTARS.2020.3021052

Appel, M., and Pebesma, E., 2019. On-demand processing of data cubes from satellite image collections with the gdalcubes Library. *Data, 4*(3), 92. https://doi.org/10.3390/data4030092

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., and Baraldi, A., 2019. Semantic Earth observation data cubes. *Data, 4*, 102. https://doi.org/10.3390/data4030102

Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., et al., 2019. Big data analytics for Earth Sciences: The EarthServer approach. *International Journal of Digital Earth, 9*(1), 3–29. https://doi.org/10.1080/17538947.2014.1003106

Cao, Q., Li, G., Yao, X., and 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

Chatenoux, B., Richard, J.-P., Small, D., Roeoesli, C., Wingate, V., Poussin, C., et al., 2021. The Swiss Data Cube, analysis ready data archive using earth observations of Switzerland. *Scientific Data, 8*, 295. https://doi.org/10.1038/s41597-021-

01076-6

Dhu, T., Giuliani, G., Juárez, J., Kavvada, A., Killough, B., Merodio, P., et al., 2019. National open data cubes and their contribution to country-level development policies and practices. *Data, 4*(4), 144. https://doi.org/10.3390/data4040144

Gao, F., Yue, P., Cao, Z., Zhao, S., Shangguan, B., Jiang, L., et al., 2022. A multi-source spatio-temporal data cube for large-scale geospatial analysis. *International Journal of Geographical Information Science, 36*(9), 1853–1884, https://doi.org/10.1080/13658816.2022.2087222

Giuliani, G., Chatenoux, B., Piller, T., Moser, F., and Lacroix, P., 2020. Data Cube on Demand (DCoD): Generating an Earth observation data cube anywhere in the world. *International Journal of Applied Earth Observation and Geoinformation, 87*, 102035. https://doi.org/10.1016/j.jag.2019.102035

Gomes, V. C. F., Queiroz, G. R., and Ferreira, K. R., 2020. An overview of platforms for big Earth observation data management and analysis. *Remote Sensing, 12*(8), 1253. https://doi.org/10.3390/rs12081253

Gomes, V. C. F., Queiroz, G. R., Ferreira, K. R., Pebesma, E., and Barbosa, C. C. F., 2024. Brazil Data Cube workflow engine: A tool for big Earth observation data processing. *International Journal of Digital Earth, 17*(1), 2313099. https://doi.org/10.1080/17538947.2024.2313099

Group on Earth Observations, 2025. Post-2025 GEO Work Programme summary document. Available online: https://earthobservations.org/storage/events/GEO-Global-Forum/geo-20/GEO-20-3.3(Rev2)_Post-2025%20GEO%20Work%20Programme.pdf (accessed on June 30, 2025).

Hoyer, S., and Hamman, J., 2017. Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software, 5*(1), 10. https://doi.org/10.5334/jors.148

Kasprzyk, J.-P., and Devillet, G., 2021. A data cube metamodel for geographic analysis involving heterogeneous dimensions. *ISPRS International Journal of Geo-Information, 10*(2), 87. https://doi.org/10.3390/ijgi10020087

Killough , B., 2018. Overview of the Open Data Cube Initiative. In *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, July 22–27, pp. 8629–8632. https://doi.org/10.1109/IGARSS.2018.8517694

Kim, K., and Lee, K., 2023. Prototyping of utilization model for KOMPSAT-3/3A Analysis Ready Data based on the Open Data Cube platform in multi-cloud computing environment: A case study. *Applied Science, 13*(18), 10478. https://doi.org/10.3390/app131810478

Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., et al., 2017. The Australian Geoscience Data Cube — Foundations and lessons learned. *Remote Sensing of Environment, 202*, 276–292. https://doi.org/10.1016/j.rse.2017.03.015

Mahecha, M. D., Gans, F., Brandt, G., Christiansen, R., Cornell, S. E., Fomferra, N., et al., 2020. Earth system data cubes unravel global multivariate dynamics. *Earth System Dynamics, 11*(1), 201–234, https://doi.org/10.5194/esd-11-201-2020

Montero, D., Kraemer, G., Anghelea, A., Aybar, C., Brandt, G., Camps-Valls, G., et al., 2024. Earth System Data Cubes: Avenues for advancing Earth system research. *Environmental Data Science, 3*, e27. https://doi.org/10.1017/eds.2024.22

Sudmanns, M., Augustin, H., Killough, B., Giuliani, G., Tiede, D., Leith, A., et al., 2023. Think global, cube local: An Earth Observation Data Cube's contribution to the Digital Earth vision. *Big Earth Data, 7*(3), 831–859. https://doi.org/10.1080/20964471.2022.2099236

Web-based European Knowledge and Earth Observation, 2023. Unlock the full potential of Earth Observation data with the CliMetLab WEkEO Plugins and xcube. Available online: https://wekeo.copernicus.eu/news/unlock-the-full-potential-of-earth-observation-data-with-the-climetlab-wekeo-plugins-and-xcube (accessed on June 30, 2025).

Yue, P., Wang, K., Xu, H., Gong, J., and Xiang, L., 2024. From Geospatial Data Cube to AI Cube: The Open Geospatial Engine (OGE) approach. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4-2024*, 441–446. https://doi.org/10.5194/isprs-annals-X-4-2024-441-2024