Ref. Ares(2024)7764014 - 31/10/2024

COMUNIDAD

D2.2
Infrastructure design document

Connecting Europe and Latin America
Transforming Today's Data into
Tomorrow's Solutions

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

comunidad-project.eu

PROJECT INFORMATION

Project Acronym	COMUNIDAD
Project Title	Combined Use of EGNSS and Copernicus Data to Develop Innovative Downstream Services for the Users from Chile and Colombia
Grant Number	101131859
Project Duration	24 months

DELIVERABLE INFORMATION

Deliverable No.	**D2.2**
Dissemination	PU
Work Package	WP2
Task	T2.2
Leading Beneficiary	Michal Kepka (BOSC)
Contributing Beneficiary (ies)	František Zadražil (Lesprojekt), Jiří Kvapil (Lesprojekt)
Peer Reviewers	Karel Charvát, Jr. (Lesprojekt), Adam Hruška (CZU)
Due date of deliverable	31.10.2024
Actual submission date	31.10.2024

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

DOCUMENT HISTORY

Version	Date	Beneficiary	Main changes
0.1	15.8.2024	BOSC	The initial version of the document
0.2	3.9.2024	LESP	Initial ArchiMate model
0.3	23.9.2024	BOSC	Section of components
0.4	8.10.2024	LESP	Finalisation of ArchiMate model
0.5	18.10.2024	BOSC	Finalisation of document for review
0.6	29.10.2024	BOSC	Processing of review comments
1.0	31.10.2024	BOSC	Final version

Table of content

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Project Summary

The COMUNIDAD project, led by Lesprojekt, utilises Copernicus satellite data and the European Global Navigation Satellite System (EGNSS), along with Artificial Intelligence (AI), Big Data technologies and numerical modelling to transfer technologies and know-how to Latin America. The COMUNIDAD project focuses on improving agricultural and forestry management in Chile and Colombia through infrastructure development and a basic platform for creating applications that enhance precision, efficiency, and sustainability. The South American region benefits from this initiative by contributing to its socio-economic growth. Technological advancements are expected to lead significantly to practical applications due to the open-source approach in development. Lesprojekt, the project coordinator, draws on its expertise in technology applications in agriculture and forestry to guide the consortium. The project provides actionable insights by employing advanced techniques to incorporate Copernicus services, EGNSS and other spatial datasets. These insights help stakeholders, including farmers, advisors, policymakers, and land managers, make informed decisions that support sustainable practices. Essential data on crop health, land use, and forestry conservation are provided, enhancing land management practices and boosting agricultural productivity.

In the COMUNIDAD project, experiences and knowledge are transferred through developing and using technological components, infrastructure, and training materials. The COMUNIDAD project aims to transform agricultural and forestry management in South America through technological innovation and international collaboration based on experiences and know-how from European partners and international cooperation with partners from Latin America. The integration of cutting-edge technologies with strategic data analysis is set to improve different domains and promote environmental sustainability in the region.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgement

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101136910.

Copyright Statement

Executive Summary

The deliverable outlines the design of the COMUNIDAD Infrastructure, which serves as a storage solution for various types of data, including thematic spatial, remote sensing, and climatic data. It features processing and analytical functions that create derivative products and publish original and processed data through standardised interfaces. These interfaces will facilitate further processing and visualisation by the COMUNIDAD Platform and user applications.

The Infrastructure was designed using the ArchiMate model described in sections 1 and 2. The detailed model is added as an image in high resolution as Appendix A. Individual components of the COMUNIDAD Infrastructure are briefly described in section 3. The types of data stored on the Infrastructure are described in section 4. The section 5 contains a description of the deployment of the COMUNIDAD Infrastructure.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Introduction

The technical solution of the COMUNIDAD project should provide functions and processes to fulfil the scope and requirements defined by general technical requirements, pilot applications requirements, and final users' needs. The COMUNIDAD Architecture will consist of the Infrastructure and the Platform where the design of the Infrastructure is the scope of the deliverable.

The COMUNIDAD Infrastructure will be the main component of the Architecture that will serve the management and integration of datasets from different sources and different scopes. It will serve as the storage and processing component that communicates with other applications through various interfaces. The target users of the Infrastructure are mainly developers of end-user applications that will use services and interfaces provided by the Infrastructure. The final users from the no-geo group will not practically notice the Infrastructure at all. The final users will interact with the Platform or end-user applications deployed on the Platform.

The design of the COMUNIDAD Infrastructure reflects requirements collected by documents "D2.1 - Requirements for infrastructure", "D4.1 - Colombian applications requirements", and the document "D4.2 - Advanced algorithm for snow coverage", which contains requirements for Chilean pilot.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

1 Design methodology

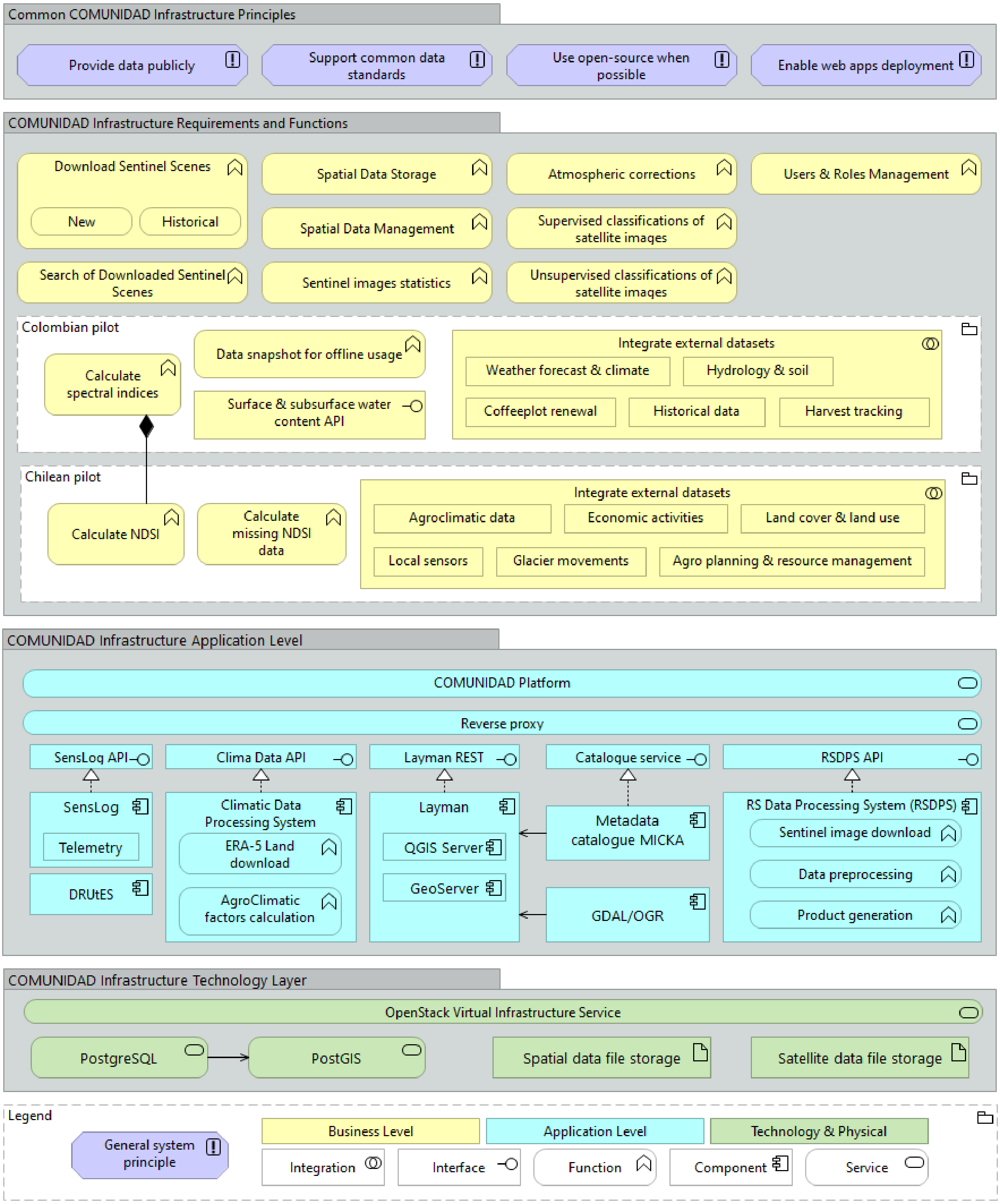
In this deliverable, we use the ArchiMate specification to model the overall system and COMUNIDAD Infrastructure, ensuring a comprehensive representation across all layers of the architecture. All schemas presented in this document adhere to the ArchiMate framework, providing a structured view that spans from high-level motivations to physical infrastructure, offering clear traceability and alignment with project goals.

The ArchiMate schema provides a standardised framework for modelling an organisation's architecture across multiple perspectives, allowing clear and structured communication between stakeholders. The schema captures elements from strategy and motivation down to infrastructure, showing the dependencies and interactions across these layers.

1. **Motivational Layer**: This layer represents the *why* of the architecture—motivating factors such as goals, principles, and drivers that inform and guide the system's design. In this schema, only general principles are used, establishing high-level goals and values without specific objectives or requirements.
2. **Business Layer**: Here, the focus is on the *what*—the business processes, functions, and roles required to meet organisational goals. The business layer outlines all necessary system functions in this context, providing a blueprint of the required capabilities and workflows.
3. **Application Layer**: This layer models the *how* of system functionality, focusing on application components and services that realise business functions. It shows how different software applications interact, the services they provide, and how they support business operations.
4. **Technology (Physical) Layer**: This final layer represents *what*, including hardware, networks, and infrastructure, support application components. It defines the physical environment where application components operate, such as servers, databases, and other technological resources.

The relationship between these layers is hierarchical and integrative: principles guide business functions; business functions drive application requirements, which then dictate physical infrastructure needs. This alignment ensures the architecture supports business objectives while aligning with core principles and technological feasibility.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.



Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

2 COMUNIDAD Architecture design

COMUNIDAD Architecture relies on the three-tier architecture, often called multi-tier architecture; it breaks down an application into distinct layers, each with a specific responsibility. This separation of concerns fosters modularity, maintainability, and scalability, making it a popular choice for web applications and complex software systems. The COMUNIDAD Infrastructure and Platform will represent this multi-tier architecture.

The three tiers that form the backbone of COMUNIDAD Architecture are the following.

1. **Presentation Tier** (User Interface, Publication layer):

This tier acts as the user's window into the application. It represents everything the user sees and interacts with, typically a web interface or a mobile app. The presentation tier collects user input through forms, buttons, and other interactive elements. It then processes this input and returns the necessary information or results to the user. The COMUNIDAD Platform will mainly provide the Presentation Tier.

2. **Business Logic Tier** (Application layer):

The business logic tier, also known as the application tier, is the heart and soul of the application. It houses the core functionalities and business rules that define the application's behaviour. This tier receives user input from the presentation tier, processes it according to the business logic, and interacts with the data tier to retrieve or manipulate data as needed. This tier will overlap the COMUNIDAD Platform and the Infrastructure. The COMUNIDAD Infrastructure will provide storage, processing, and publishing by APIs. At the same time, the Platform will utilise services provided by the COMUNIDAD Infrastructure and further process data to provide the final products to users.

3. **Data Tier:**

The data tier, or the persistence tier, is responsible for storing and managing the application's data. It is the permanent repository for data accessed by the business logic tier. Databases are the main components of this tier. They ensure data integrity, security, and efficient retrieval when needed by the business logic. The data tier will be the main component of the Infrastructure; it will store original or processed data that will be published by different APIs for further processing by the platform.

The separation of concerns in three-tier architecture offers several advantages:

- Modularity: Each tier performs a specific function, making it easier to develop,

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

maintain, and modify individual components independently.

- Scalability: Different tiers can be scaled independently based on their specific needs. For instance, you can scale the presentation tier to handle more users while scaling the database server to accommodate growing data volumes.
- Maintainability: Fixing bugs or adding new functionalities becomes easier as developers can focus on a specific tier without worrying about the intricate workings of the entire system.
- Security: The three-tier architecture enhances data security by placing the data tier behind a secure layer.
- Reusability: The business logic and data access layers can often be reused across different applications, saving development time and resources.

3 COMUNIDAD Infrastructure components

This chapter describes in detail the individual components of the COMUNIDAD Infrastructure and their interfaces for processing components.

3.1 Open-Source Powerhouse (OpenStack)

Cloud solution powering COMUNIDAD Infrastructure leverages the power of OpenStack, a leading open-source cloud software platform. This technology efficiently utilises hardware resources by seamlessly combining individual servers into a unified, centrally managed cloud environment. OpenStack empowers COMUNIDAD Infrastructure with a comprehensive suite of services, including:

- Virtualisation: Create and manage virtual machines (VMs) on demand, offering a cost-effective and flexible way to run diverse applications.
- Block Device Sharing: Efficiently share storage resources across VMs, ensuring data accessibility while optimising storage utilisation.
- Distributed Computing: Harness the combined processing power of multiple servers for computationally intensive tasks, significantly accelerating workloads.
- Docking Technology Support: Integrate containerisation technologies like Docker for simplified application deployment and management.
- Advanced Networking: Benefit from a robust network stack that enables low-level network isolation, including VLAN support, Network Address Translation (NAT), and

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

an integrated firewall, guaranteeing secure and controlled network communication.

Implement fine-grained user access controls to logically segment OpenStack instances into isolated environments, replicating the functionality of separate physical servers. OpenStack prioritises user experience. Through a comprehensive web dashboard, administrators gain a clear and detailed overview of current resource utilisation, enabling proactive management and resource allocation. Additionally, robust reporting capabilities offer valuable insights into historical resource usage patterns.

Data security is paramount. Backups are conducted at least weekly and more frequently for mission-critical infrastructure. Lesprojekt manages its dedicated backup infrastructure, providing complete control and guaranteeing data restoration capabilities in unforeseen circumstances.

COMUNIDAD Infrastructure caters to evolving needs. New virtual servers can be provisioned within minutes in response to user requests, eliminating delays and ensuring rapid application deployment. A library of pre-configured operating system snapshots is readily available for immediate use, further streamlining the deployment process.

3.2 Reverse Proxy

The entire COMUNIDAD Architecture resides within a Network Address Translation (NAT) environment, protected by a reverse proxy server for increased security, manageability, and scalability. This reverse proxy is a single-entry point for all user requests directed towards the COMUNIDAD domain.

There are several benefits of the Reverse Proxy. The reverse proxy shields the internal components of the COMUNIDAD Architecture from direct internet access, adding an extra layer of protection.

The proxy manages incoming requests, directing them to the appropriate backend server based on the specific path within the URL (URI path). This simplifies system administration by presenting a unified address despite the heterogeneous nature of the backend servers (e.g., portal, map server). The reverse proxy can be easily configured to handle increased traffic volumes, ensuring the COMUNIDAD solution scales efficiently as user demands grow.

In essence, the reverse proxy acts as a central traffic controller, ensuring a secure and efficient user experience while simplifying the management of the underlying

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

infrastructure.

3.3 PostgreSQL database

PostgreSQL, a widely adopted open-source relational database management system (RDBMS), is a robust foundation for managing various data types. PostgreSQL transforms into a comprehensive Geographic Information System (GIS) database with the PostGIS extension. PostGIS unlocks a suite of functionalities specifically designed for storing, manipulating, and analysing geospatial data.

The integration of PostGIS empowers users to represent diverse geographic features – points, lines, polygons, and even 3D geometries – directly within the PostgreSQL database. This seamless integration eliminates the need for separate storage solutions for geospatial data, fostering a centralised and efficient data management environment.

PostGIS boasts a rich set of spatial functions beyond primary storage and retrieval. These functions enable complex spatial operations such as calculating distances and areas and performing advanced geometric analyses. These analytical capabilities empower researchers and data scientists to extract valuable information from their geospatial datasets.

Spatial indexing, a cornerstone of PostGIS functionality, significantly enhances the efficiency of retrieving specific locations based on geospatial queries. This indexing mechanism allows for rapid filtering and retrieval of relevant data, making it a crucial factor for applications that demand real-time geospatial information.

The open-source nature of PostgreSQL and PostGIS fosters a vibrant community that contributes to extensive documentation and a wealth of resources. This robust support system empowers users of all levels to leverage the power of geospatial data management within their PostgreSQL databases.

PostgreSQL's robust data management capabilities and PostGIS's specialised functionalities create a powerful duo for handling geospatial data. This integrated solution streamlines data management, facilitates efficient retrieval, and empowers in-depth spatial analysis, paving the way for informed decision-making across various disciplines.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

3.4 Metadata Catalogue (MICKA)

MICKA is a comprehensive system for managing and publishing spatial data metadata. It is built on the ISO 191xx and INSPIRE standards and supports the OGC CSW 2.0.2 ISO AP-1.0 catalogue service. Also, MICKA offers various advanced features and supports multiple other standards, ensuring interoperability with other applications. MICKA Key Features:

- Online Metadata Editing: MICKA provides a user-friendly web application for online metadata editing.
- CSW 2.0.2 ISO-AP Support: The system fully supports CSW 2.0.2 ISO-AP, including transactions and harvesting.
- Feature Catalogue Management: MICKA enables efficient management of feature catalogues according to ISO 19110.
- Dublin Core Metadata Support: The system supports metadata based on the Dublin Core standard (ISO 15836).
- Import from OGC Services: MICKA facilitates metadata import from various OGC services, including WMS, WFS, CSW, and WCS.
- Metadata Import from ISO 19139, CKAN, and Others: The system supports metadata import from various formats, including ISO 19139, CKAN, and others.
- Multilingual Metadata: MICKA allows for the creation and management of multilingual metadata.
- External Registry and Thesaurus Integration: The system can integrate external registries and thesauri, such as INSPIRE registries, GEMET, and SPARQL.
- Configurable User Vocabularies: MICKA allows for the configuration of user-defined vocabularies.
- INSPIRE Atom Download Service: The system offers the option to create an INSPIRE Atom download service.
- Built-in Metadata Validator: MICKA features a built-in metadata validator to configure multiple profiles.

MICKA also includes a metadata validator that can be used to validate spatial data metadata and OGC services against INSPIRE requirements and the national metadata profile of the Czech Republic. The validator can be used as a standalone application or service implemented in the National INSPIRE Geoportal of the Czech Republic. It is also available as part of the MICKA metadata catalogue.

MICKA and its associated catalogue client and metadata validator provide a

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

comprehensive suite of tools for effectively managing, publishing, and validating spatial data metadata. The system's support for various standards, user-friendly interface, and extensive customisation options make it a valuable solution for organisations and individuals working with spatial data.

3.5 Layman

Layman facilitates the online publication of geospatial data through a REST API. It offers two publication paradigms: layers and maps. Layers represent individual vector or raster datasets, including time series for raster data. Maps are collections of these layers. Layman exhibits broad data format support, accepting vector layer data in GeoJSON, Shapefile, or directly from PostGIS tables referenced by PostgreSQL connection URIs. Similarly, raster data can be uploaded in GeoTIFF, JPEG2000, PNG, and JPEG formats. Additionally, Layman supports styling for vector and raster data via Styled Layer Descriptor (SLD) and QGIS Style File Format. Map definitions are specified in the HSLayers Map Composition format.

Asynchronous chunk upload enables seamless upload of large files from web browsers. Layman utilises OAuth2 for user authentication and enforces authorisation by granting specific users read/write access to individual layers and maps. Processing tasks are handled asynchronously to enhance performance.

Layman offers a comprehensive suite of services through URL endpoints, including a REST API, Web Map Service (WMS), Web Feature Service (WFS), and Catalogue Service (CSW). The platform prioritises security, providing documented details on its security system, data storage practices, and proxying mechanisms. Layman leverages environment variables for configuration flexibility.

3.6 Remote Sensing Data Processing System (RSDPS)

The Remote Sensing Data Processing System (RSDPS) is a robust pipeline designed to efficiently handle satellite imagery, from initial download to final product dissemination. The system automates the intricate process of acquiring, processing, storing, and publishing satellite data, primarily focusing on Sentinel-1 and Sentinel-2 imagery sourced from the Copernicus Data Spaces Environment.

It uses Python and GDAL-based tools as its main components. MAJA is used for atmospheric corrections, and SNAP carries out advanced processing and analysis of

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

radar data. Core Functionalities:

- **Data Acquisition**:
  - Seamlessly downloads raw Sentinel-1 and Sentinel-2 satellite scenes.
  - Extract essential metadata (date, scene nomenclature, cloud coverage, etc.) for efficient organisation and future reference.
  - Populates the metadata catalogue with extracted information.
- **Pre-processing:**
  - Tailored pre-processing steps are applied based on image type (optical or radar).
  - Raw images are rigorously corrected to mitigate sensor bias and atmospheric interference.
  - This critical step ensures the consistency and comparability of different scenes, enabling accurate analysis and monitoring over time.
- **Product Generation:**
  - A diverse range of products can be derived from the pre-processed imagery, including Vegetation indices and Supervised classifications (leveraging machine learning techniques like artificial neural networks)
  - These products offer valuable insights into various environmental parameters, such as vegetation health, forest cover, soil moisture, and deforestation patterns.
  - The specific product suite will be determined in collaboration with stakeholders and pilot areas (customised set of products and classification schemas, local coordinate system)
- **Data Storage and Dissemination**:
  - Processed data and source imagery are securely stored in a centralised repository, ensuring easy accessibility and long-term preservation.
  - A user-friendly web interface facilitates data searching, downloading, and visualisation.
  - Web services enable integration with third-party systems and mapping applications, empowering users to leverage the data for their specific needs.

The RSDPS offers a comprehensive solution for managing and analysing satellite imagery. Automating complex processing tasks and providing a user-friendly interface empowers researchers, policymakers, and other stakeholders to extract valuable information from Earth observation data.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

3.7 Climatic Data Processing System (CDPS)

The Climatic Data Processing System is a processing application designed to use the ERA5-Land dataset from the Copernicus Climate Change Service (C3S). The system automates the intricate process of acquiring, processing, storing, and publishing climatic data, mainly the ERA5-Land dataset. ERA5-Land is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades at an enhanced resolution compared to ERA5.

The CDPS uses Python and Jupyter Notebooks tools as its main components. It obtains NetCDF files from the C3S for a defined period and area. It stores original data in the database and processes different algorithms to calculate agro-climatic factors.

The calculated agro-climatic factors are frost dates, frost-free periods, growing degree units, heat stress units, number of growing days, number of optimal growing days, dates of fall nitrogen application, precipitation, evapotranspiration, and runoff sums together as water balance and solar radiation.

The output of the processing system is API, which provides GeoJSON/Shapefiles with original data or calculated agroclimatic factors.

3.8 SensLog

SensLog is an open sensor data management solution for receiving, storing, managing, analysing, and publishing sensor data. This solution is suitable for static in-situ sensors, sensors on mobile carriers, and Volunteered Geographic Information (VGI) gathered by smart devices. SensLog is a backend solution that runs on a web server and stores data in an RDBMS. SensLog is a modular solution where every module is designed as a separate component. Thus, the scalability of the solution is ensured.

SensLog stores data in a relational data model based on the ISO 19156 Observations & Measurements standard, but it contains additional extensions mainly for user management and user hierarchy and for the hierarchy of the sensor network. Different profiles for sensor tracking, VGI or complex observations extend the core of the model that stores numerical observations. The data model is implemented in PostgreSQL RDBMS with the spatial extension PostGIS.

SensLog provides different interfaces, but the REST API with JSON data encoding is used mainly. The REST API offers proprietary JSON formats that cover the whole functionality of the solution. Sensor data and observation publication is possible by services defined

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

by the OGC SOS v1.0 standard. The observation inserting and publishing are also manageable by the OGC SensorThingsAPI 1.0 standard. The VGI data can be exported in RDF format or as binary data with multimedia.
The interoperability of SensLog is covered by a system of SensLog Connectors. SensLog Connector is a component that translates one interface to a different one. SensLog Connector can be used in 3 versions – push, pull or proxy. Push and pull versions have SensLog REST API on the input or output side, and the data provider or consumer completely defines the second. The proxy version has external APIs on both the input and output sides.

3.9 GDAL/OGR

GDAL and OGR are important libraries for data processing. Geospatial Data Abstraction Library (GDAL) is a translator library for raster and vector geospatial data formats. The related OGR library (OGR Simple Features Library), part of the GDAL source tree, provides a similar ability for simple features vector graphics data. The library supports more than 150 formats for raster and 90 vector formats.

3.10 DRUtES

DRUtES is a finite element method solver specifically designed to address the Richards equation problem. In the COMUNIDAD project, DRUtES will integrate ERA5 data series, soil hydraulic properties, and the root water demand of specific crops to assess soil moisture comfort for these crops.
The soil hydraulic properties and root water demand parameters are already available for the Colombian pilot site and could be easily extended to larger areas based on local expert knowledge.
Within the proposed COMUNIDAD Infrastructure, DRUtES will be executed on a daily basis, beginning with a model warming period of five months. The initial condition will assume average soil water saturation. It is expected that the five-month model warming period will establish an initial-condition-independent setup; however, if needed, this period could be extended. Since there is always a data gap between the current day and the ERA5 historical data, this gap will be filled by meteorological nowcast data obtained from a selected meteorological provider, such as the global Norwegian Meteorological Institute forecast (https://yr.no), which offers an API interface, or, if available, a local

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

weather forecast provider.
With DRUtES, we can provide the current state of soil moisture comfort, as well as an antecedent trend covering the previous 14 days and, if requested, a short-term forecast trend of up to seven days. This component will rely exclusively on the numerical modelling of soil moisture dynamics. We will not utilise soil moisture data from satellite images, as interpreting soil moisture data from satellite images involves uncertainties that cannot be easily resolved at the project's pilot site locations.

- **Input Data**: ERA5 data series for a specified location and time frame, covering the previous six months.
- **Output**: Computed soil moisture and categorised soil moisture comfort levels, classified as *overflooded, oversaturated, optimal, dry,* and *overly dry*. This output covers a period of the previous 14 days with an hourly time step. Data will initially be distributed in CSV file(s). If more spatial heterogeneity is considered, data could also be distributed as NetCDF files.

3.11 Distributed Computing in OpenStack

Distributed Computing in OpenStack enables leveraging the combined processing power of multiple servers to tackle computationally intensive tasks. This approach significantly accelerates workloads by dividing complex processes across various nodes in a scalable, cloud-based environment. Within the COMUNIDAD project, distributed computing is instrumental in efficiently calculating various indices and computer vision algorithms, essential for large-scale data analysis and real-time processing.
This distributed setup is designed to handle diverse processing demands. For instance, calculations of vegetation indices or environmental metrics can be parallelised, allowing for rapid data analysis across extensive datasets. Similarly, compute-intensive computer vision algorithms, often used for object recognition or geospatial analysis, can be divided across the distributed nodes, enhancing performance and reducing computation times.
OpenStack provides a flexible environment for distributed computing with robust support for load balancing, fault tolerance, and resource scaling. Its architecture enables the creation of virtual machines or containers across multiple physical servers, ensuring computational resources are dynamically allocated based on task demands. With this scalable infrastructure, distributed computing becomes a powerful tool within

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

OpenStack, enabling real-time insights and processing capabilities essential for advanced data analysis and computationally demanding algorithms.

4 COMUNIDAD Infrastructure data content

Many types of data from different providers will be stored and published by the COMUNIDAD Infrastructure. This section describes the most common data types stored on the COMUNIDAD Infrastructure.

4.1 Spatial data

Most of the data content of the COMUNIDAD Infrastructure can be considered spatial data, but this section will cover thematic spatial data not mentioned in further sections. Besides the global harmonised datasets, there are many national datasets suited to the needs of a given pilot country. These play a crucial role for pilot localities and will be used by original services of data providers or integrated into the COMUNIDAD Infrastructure as the central point for further utilisation by the applications deployed on the COMUNIDAD Platform. It is important to mention that many spatial datasets are provided in individual countries' legal geodetic reference systems. To ensure the interoperability and reuse of data, the COMUNIDAD Infrastructure could transform global data to selected national grids of pilot countries besides the original coordinate reference system (CRS).

4.2 Satellite images and products

RSDPS will manage the download and processing of the Sentinel-1 and Sentinel-2 data, which will be further processed to generate products like indices or classifications. The original data will be provided for further processing by applications deployed on the COMUNIDAD Platform. The GeoServer will publish satellite images and derived products using standardised web services (WMS, WMTS, WCS).

4.3 Climatic data

The CDPS will manage the ERA5-Land data from the Copernicus Climate Change Service (C3S) for selected pilot localities and defined periods, which will grow during the project period. Besides the original ERA5-Land snapshot, the CDPS will publish the location-based agroclimatic factors for selected pilot localities. Climatic data and calculated factors will be published using REST API with JSON or GeoJSON encoding.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

4.4 Metadata

Metadata records are planned for each integrated dataset or service provided by components of the COMUNIDAD Infrastructure. Metadata will be managed and published by the MICKA metadata catalogue.

5 COMUNIDAD Infrastructure deployment

The COMUNIDAD Infrastructure is intended to run on an internal cloud solution managed by Lesprojekt that offers a robust and customisable environment specifically designed to support the development and deployment of services for the COMUNIDAD Infrastructure and the Platform. The whole infrastructure is built on OpenStack software, based on Linux operating systems and other open-source technologies.

OpenStack enables optimal usage of hardware resources, joining standalone servers in a seamless cloud under a single management. It provides many services, including virtualisation, block device sharing, distributed computing, support for various docking technologies, etc. A robust network stack allows for low-level network isolation, including VLAN support, NAT and an integrated firewall. Fine-grained user management enables logical segmentation of the OpenStack instance into standalone, independent instances as if run on separate physical servers.

The administrator has an excellent overview of the current use of hardware resources due to solid reporting capabilities and a web dashboard.

Backup is done at least weekly, for critical infrastructure even more frequently and stored at geographically different localities. Lesprojekt manages the company's own backup infrastructure.

New virtual servers are rapidly deployed within minutes, as our users request. Various operating systems are prepared as snapshots for instant use.

Built on a dedicated server infrastructure located in Prague, Czech Republic, the cloud solution provides complete control and ensures the utmost security for COMUNIDAD data and applications.

The COMUNIDAD Infrastructure design is focused on future transferability to the pilot countries and maintaining the COMUNIDAD Infrastructure there on their local servers.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Conclusion

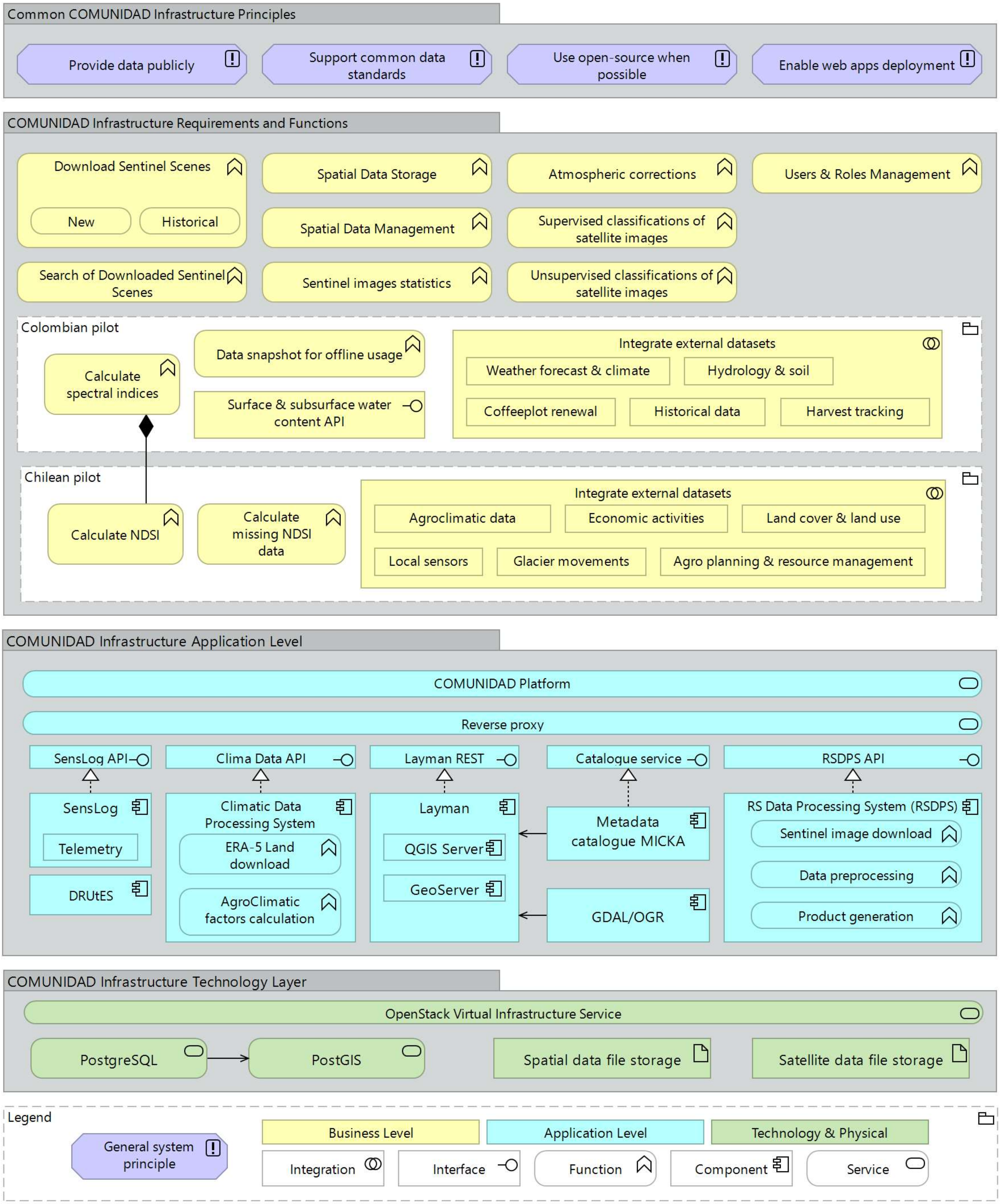
The deliverable describes the design of the crucial component of the COMUNIDAD Architecture - Infrastructure. The COMUNIDAD Infrastructure provides storage for various types of data, especially thematic spatial data from global or local producers or providers, remote sensing data and climatic data. Processing and analytical functions and applications will process stored data to various derivate products and publish original and processed data by different interfaces. The COMUNIDAD Infrastructure will provide standardised interfaces where possible (e.g. OGC-defined web services) or interfaces according to common practice in the domain. The COMUNIDAD Infrastructure interfaces will be the main point for using the Infrastructure for further processing or visualisation by the COMUNIDAD Platform itself and user applications deployed on the Platform.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.

Appendix A.

The following image presents a detailed infrastructure model according to the ArchiMate specification.

Funded by the European Union under grant agreement No. 101131859. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUSPA. Neither the European Union nor the granting authority can be held responsible for them.


Common COMUNIDAD Infrastructure Principles
Provide data publicly
Support common data standards
Use open-source when possible
Enable web apps deployment
COMUNIDAD Infrastructure Requirements and Functions
Download Sentinel Scenes
New
Historical
Spatial Data Storage
Atmospheric corrections
Users & Roles Management
Spatial Data Management
Supervised classifications of satellite images
Search of Downloaded Sentinel Scenes
Sentinel images statistics
Unsupervised classifications of satellite images
Colombian pilot
Calculate spectral indices
Data snapshot for offline usage
Surface & subsurface water content API
Integrate external datasets
Weather forecast & climate
Hydrology & soil
Coffeeplot renewal
Historical data
Harvest tracking
Chilean pilot
Calculate NDSI
Calculate missing NDSI data
Integrate external datasets
Agroclimatic data
Economic activities
Land cover & land use
Local sensors
Glacier movements
Agro planning & resource management
COMUNIDAD Infrastructure Application Level
COMUNIDAD Platform
Reverse proxy
SensLog API
Clima Data API
Layman REST
Catalogue service
RSDPS API
SensLog
Telemetry
DRUtES
Climatic Data Processing System
ERA-5 Land download
AgroClimatic factors calculation
Layman
QGIS Server
GeoServer
Metadata catalogue MICKA
GDAL/OGR
RS Data Processing System (RSDPS)
Sentinel image download
Data preprocessing
Product generation
COMUNIDAD Infrastructure Technology Layer
OpenStack Virtual Infrastructure Service
PostgreSQL
PostGIS
Spatial data file storage
Satellite data file storage
Legend
General system principle
Business Level
Application Level
Technology & Physical
Integration
Interface
Function
Component
Service