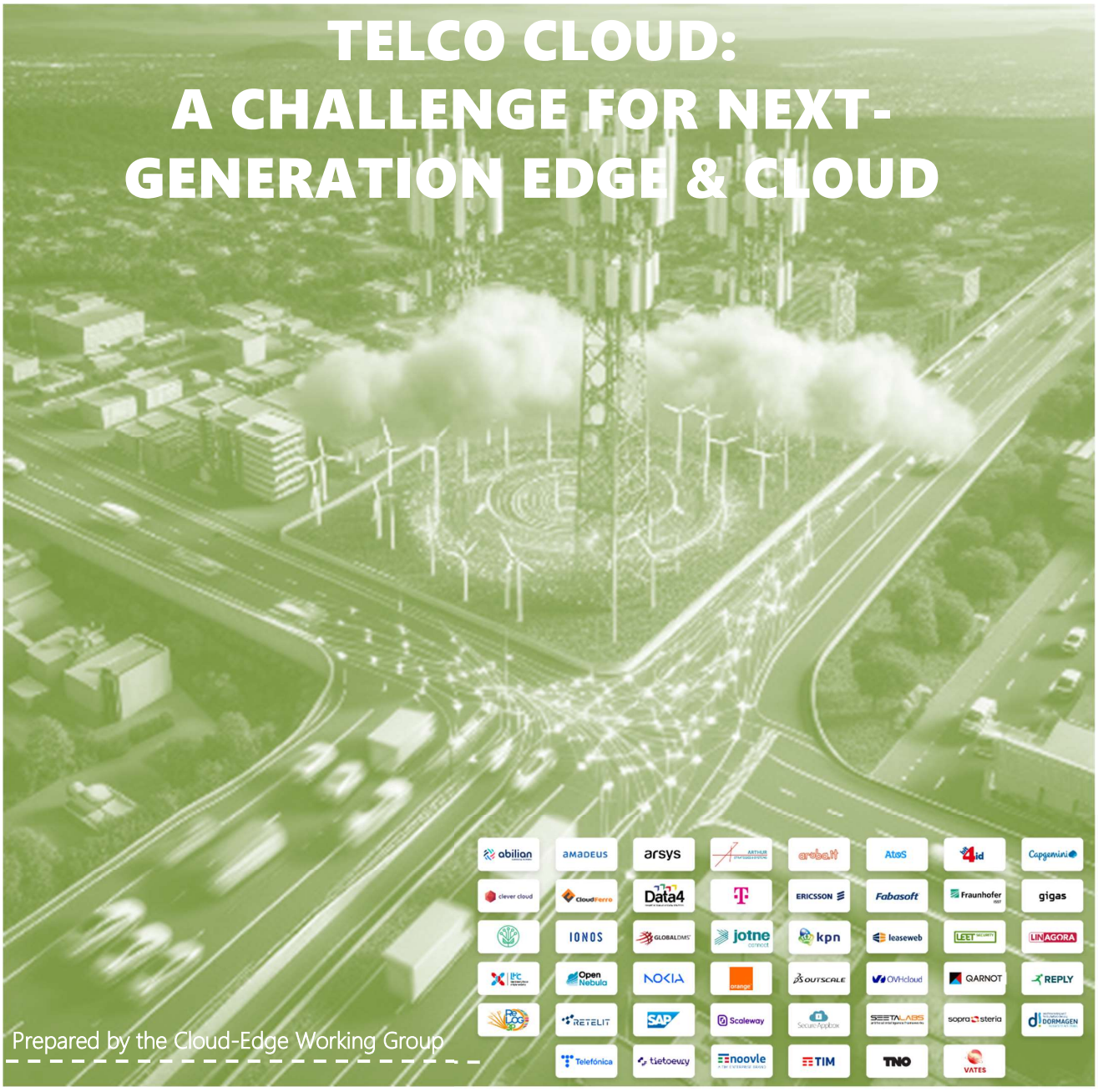
EUROPEAN ALLIANCE
FOR INDUSTRIAL DATA,
EDGE AND CLOUD

THEMATIC ROADMAP

TELCO CLOUD: A CHALLENGE FOR NEXT-GENERATION EDGE & CLOUD



Prepared by the Cloud-Edge Working Group

MAIN CONTRIBUTORS

Alfonso Carrillo Aspiazu (OpenNebula Systems), Alessandro Capello (TIM), Juan Carlos Garcia (Telefónica), Andrea Calvi (TIM), Carlo Cavazzoni (TIM), Andreas Florath (Deutsche Telekom), Edwin Harmsma (TNO), Nicolas Homo (Orange), Borgert van der Kluit (TNO), Norbert Niebert (Ericsson), Charles Schulz (Vates), Luis Velarde (Telefónica), Arthur van der Wees (Arthur's Legal, Strategies & Systems).

EDITORS

Juan Carlos García (Telefónica), Dimosthenis Kyriazis (University of Piraeus)

ABOUT THE ALLIANCE & THE WORKING GROUP

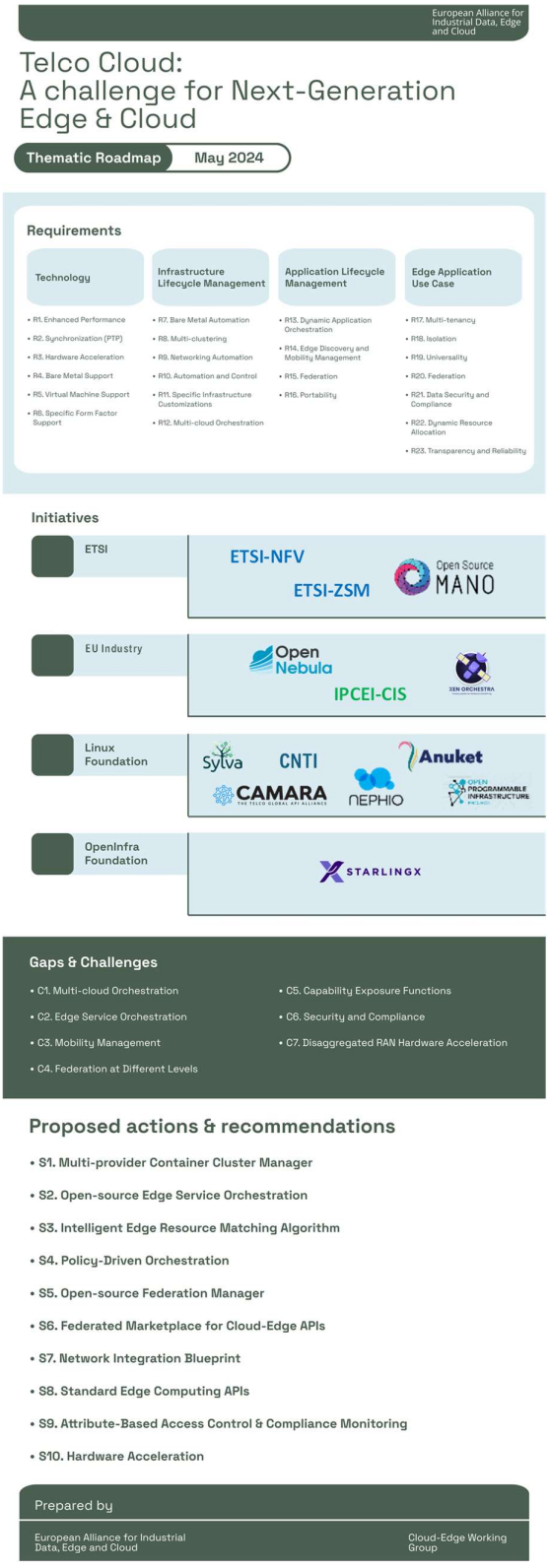
The **European Alliance for Industrial Data, Edge and Cloud** [1] brings together businesses, Member States' representatives, and relevant experts to jointly define strategic investment roadmaps to enable the next generation of highly secure, distributed, interoperable, and resource-efficient computing technologies. The work is facilitated by the European Commission's Directorate-General for Communications Networks, Content and Technology (DG CNECT).

The **Cloud - Edge Working Group** (WG) brings together the main European Industry players in cloud computing that compiled the European Industrial Technology Roadmap for the Next-generation Cloud-Edge in 2023 [2]. Following the release of the aforementioned roadmap, specific WG members have collaborated on developing the current thematic roadmap with a focus on cloud environments for the telecommunications industry (referred to as "Telco Cloud") as a challenge and enabler for next-generation Cloud-Edge. The members of the WG that have co-authored this roadmap are the followin:

Arthur's Legal, Strategies and Systems	OpenNebula Systems SL
Capgemini	Orange
Deutsche Telekom	Telefonica
Ericsson AB	TIM SpA
Nokia	TNO

The WG is chaired by Ignacio M. Llorente (OpenNebula Systems), co-chaired by Jean-Philippe Defrance (Capgemini) and Mark Kuehner (SAP), and facilitated by Ana Juan Ferrer (European Commission). The group that developed the current thematic roadmap is chaired by Juan Carlos García (Telefónica).

EXECUTIVE SUMMARY



The current thematic roadmap, entitled **Telco Cloud: A Challenge for Next-Generation Edge & Cloud**, focuses on challenges and recommendations emerging from an analysis of the current landscape of the telecommunications industry and the interplay with emerging edge and cloud computing environments. The roadmap covers initiatives (as analysis of the baseline), gaps, and challenges in order to conclude with a set of recommendations to realise the vision of a telco cloud.

In this context, the current thematic roadmap introduces a structure that provides information on the current landscape in terms of requirements defining the corresponding need, relevant existing (or ongoing) research and commercial solutions, the gaps emerging following the the mapping between requirements and existing solutions, and the recommendations to tackle the aforementioned gaps. To this end, the main pillars tackled by this roadmap are the following:

- Requirements, including technology, infrastructure and application lifecycle management as well as relevant edge use cases.
- Initiatives providing existing research and commercial solutions, including standardisation organisations, EU-funded projects and open source projects.
- Gaps & challenges following the analysis of the requirements and the existing solutions in several technological areas such as federation, mobility, orchestration, security, and hardware acceleration.
- Recommendations in the scope of projects, policies, and funding programmes.

TABLE OF CONTENTS

About the Alliance & the Working Group **3**

Executive Summary **4**

Introduction **7**

Section 1: Telco Requirements for the Cloud-Edge Continuum **9**

2.1 Telco Use Case 9
2.2 Edge Application Use-Case 14

Section 2: Initiatives addressing Telco Requirements **16**

2.1 OpenNebula 16
2.2 ANUKET 17
2.3 SYLVA 17
2.4 NEPHIO 19
2.5 CAMARA 20
2.6 OPI 21
2.7 StarlingX 22
2.8 ETSI-NFV 22
2.9 ETSI-OSM 25
2.10 ETSI-ZSM 26
2.11 IPCEI-CIS 27
2.12 Vates VMS 28
2.13 Other related activities 29

Section 3: Challenges and Requirements to be addressed **30**

C1. Multi-cloud Orchestration 30
C2. Edge Service Orchestration 30
C3. Mobility Management 31
C4. Federation at Different Levels 32
C5. Capability Exposure Functions 33
C6. Security and Compliance 34
C7. Disaggregated RAN hardware acceleration 34

Section 4: Proposed Actions and Recommendations **36**

S1. Multi-provider Container Cluster Manager 36

S2. Open source Edge Service Orchestration 36
S3. Intelligent Edge Resource Matching Algorithm 36
S4. Policy-Driven Orchestration 37
S5. Open source Federation Manager 37
S6. Federated Marketplace for Cloud-Edge APIs 37
S7. Network Integration Blueprint 38
S8. Standard Edge Computing APIs 38
S9. Attribute-Based Access Control & Compliance Monitoring 38
S10. Hardware Acceleration 38

Section 5: Future Topics 40

Section 6: Conclusions 41

References 42

INTRODUCTION

At an ever-increasing pace, technology is constantly shaping society and influencing every sector of our local, regional, national, and global economies, providing the impetus needed to accelerate digital transformation and to address various societal challenges expected in the mid- and long-term. Digital infrastructures and solutions are crucial for all economic and public systems as they represent an extremely effective suite of technologies and tools for improving productivity and sustainability, while at the same time meeting citizens' needs.

As per the EU Fit for the Digital Age mission [3], and in order to deliver the so-called "Path to the Digital Decade 2030" [4], a set of objectives and targets has been defined. One of these targets refers to the deployment of 10,000 climate-neutral, highly secure edge nodes in the EU by 2030 - in a way that will guarantee access to computing and data services with low latency wherever people and organisations are located.

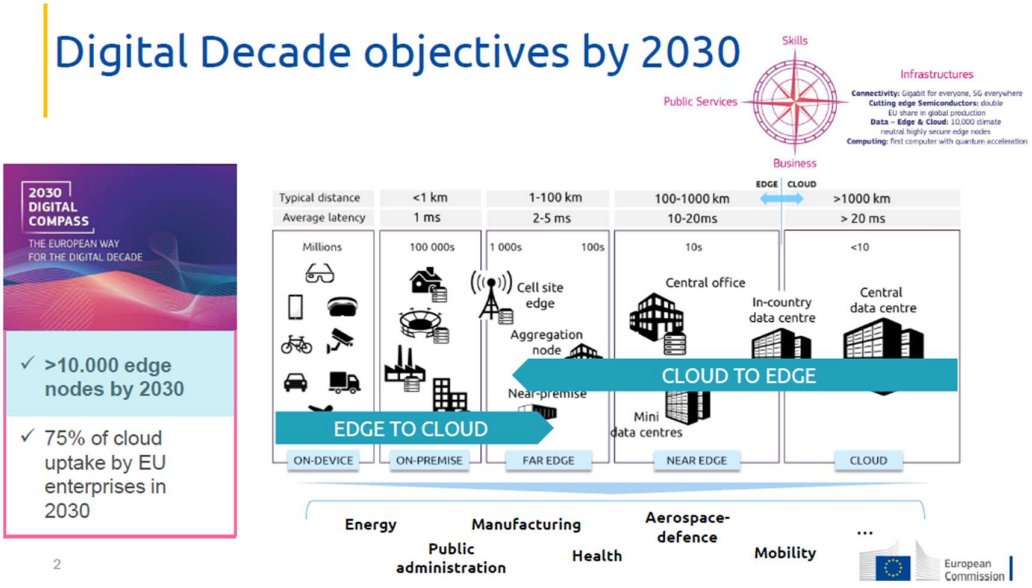


Figure 1: EU's Digital Decade objectives [4].

Following the transformation in other industries, the evolution of network systems towards cloud-native implementations is happening in all network domains, thus demanding extensive computing capacity at multiple edge and cloud locations. Thus, telecom operators will be one of the relevant tenants for the cloud-edge continuum in the short-medium term.

In the short-term, the 5G SA Core represents the first cloud-native 3GPP mobile technology. 5G SA Core is in the commercial phase and is being deployed by most of the operators worldwide. In the

medium-term (2025-30), there will also be radio access (RAN), fixed access (FTTH/PON), and transport networks implemented as cloud-native solutions.

All these network domains will need a highly distributed computing infrastructure, from device and far edge to centralised data centres, to deploy and execute different network functions. Some network domains may have specific requirements from the underlying cloud infrastructure that provides their runtime environment. These requirements are associated to several aspects:

- Specific hardware configurations (accelerators, NICs, etc.), to ensure proper performance and efficiency of the solution.
- Operation support technologies to manage the distributed computing infrastructure and the chains of network functions deployed on top.
- Automation and orchestration approaches to manage the lifecycle of the vast and highly interlinked set of applications (network functions) that are required to deliver a network service.

Addressing these topics will bring significant benefits to the future European cloud-edge continuum, as further detailed in the next sections of the current roadmap. Moreover, by addressing them for a particularly demanding and complex use case, like a telecommunication service, it will provide capabilities that will also bring value to other types of applications. Telecom networks will demand a high amount of computing capacity, which will realise important scaling to any edge and cloud environment hosting them, being one of the opportunities for the European cloud to grow.

This thematic roadmap complements the work done by the Edge and Cloud WG as outlined in its "*European Industrial Technology Roadmap for the Next-Generation Cloud-Edge*" [2], handed over to the Commission in July 2023 ("Roadmap"), by:

- Further developing one of the focus areas, namely "*Achieving scale at the Edge by hosting Telco Network Functions*" covered in detail in Chapter 7 "*A New Connectivity for the Edge and Cloud*" of the Roadmap.
- Introducing aspects of "*Cloud - Edge Foundation Infrastructure*" (chapter 8 of the Roadmap) in relation to the specific infrastructure requirements.
- Tackling "*Interoperability and Multi-Provider Services*" (as set forth in Chapter 5 of the Roadmap) regarding portability in digital ecosystems and the fact that the network environment will be hybrid and multi-cloud.
- Highlighting specific aspects relevant to network systems with regards to digital sovereignty, sustainability, and security (Chapters 2, 3, and 4 of the Roadmap).

SECTION 1: TELCO REQUIREMENTS FOR THE CLOUD-EDGE CONTINUUM

Telecom operators are positioned to play a double role in the evolving cloud and edge computing landscapes. On one side, internal network enhancements are propelled by a continual shift towards more flexible and adaptable technologies. This requires a wide array of hardware and software configurations, operation support tools, and automation mechanisms to enable a highly available cloud environment for public networks[1]. On the other side, sharing the computing infrastructure at edge locations for telecom network functions and external workloads presents both a unique opportunity to increase resource utilisation as well as posing a challenge. This duality depicts an opportunity both for telecom services and for customer applications to contribute to the scale of the cloud-edge infrastructure.

This section provides a thorough examination of two key meta-use-cases, emphasising their relevance for the European Commission's Digital Strategy. Proposed initiatives designed to capitalise on these opportunities are also presented. Interoperability, sustainability, and security are addressed, aligning with existing European technology roadmaps.

2.1 Telco Use Case

Cloud technologies bring flexibility, scalability, and adaptability to the telecom environment, and increase its availability and durability. The use of open source technologies such as Kubernetes (for container management) allows a **disaggregated approach** that facilitates the deployment, interoperability, reversibility, scalability, and evolution of software solutions.

However, the adoption of a cloud-native architecture implies additional complexity and risks that need to be considered. The complexity emerges from additional **abstraction and control layers** required for flexible and efficient compute utilisation**, network disaggregation** enabling innovative and cost-effective solutions, new **orchestration** components for managing a heterogeneous runtime environment, the need for **AI/ML** to improve optimisation of assets and process and closed-loop automation, **exposure** mechanisms to monetise telco capabilities, etc. The risks emerge, among other reasons, from the fact that networks move from proprietary appliances to open and general-purpose hardware and mainstream cloud technologies. On the other hand, it should be noted that the exploitation of such technologies brings extensive expertise and tools to alleviate complexity and risk.

[1] A *public network* provides service to multiple customers who share the same physical infrastructure to receive the service, as opposed to private networks that serve a single customer with dedicated physical (or virtual) resources.

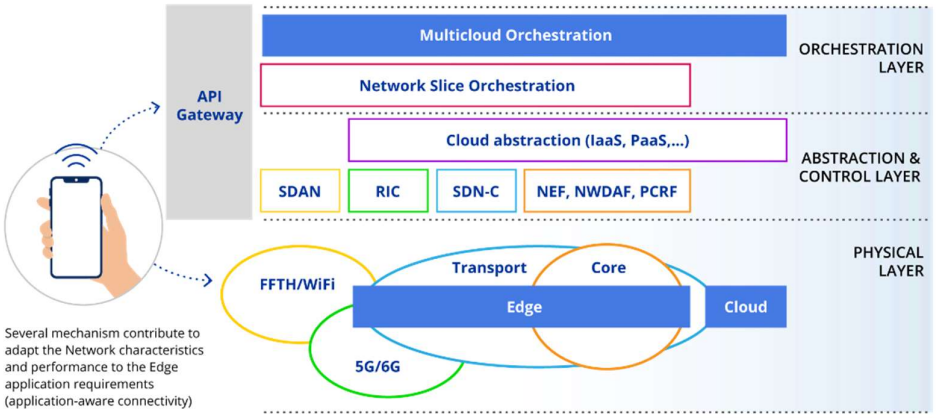


Figure 2: Additional abstraction, control, orchestration, and exposure layers in a cloud-native telco infrastructure.

Several challenges are intrinsic to implementing internal use-cases in telecom services. These encompass a wide range of issues, from the need for distributed computing architectures to specific operational toolsets. The complexity of deploying and managing internal network functions often places constraints on resources like space, power, or environmental conditions, and poses risks such as potential lack of compatibility between solutions that require extraordinary integration, test, and deployment efforts if not adequately addressed.

Complexity increases as new requirements like latency, jitter, or upstream bandwidth are demanded by the new 5G and Beyond 5G applications and use cases, resulting in traffic patterns that are less deterministic and change more dynamically. Data-driven and AI/ML-assisted learning and decision making will be needed to quickly adapt the network topology, capacity, and configuration to this dynamic environment.

Telcos need to deploy *hundreds of edge cloud nodes* in the coming years, accommodating in the same location telco functionality from different network domains such as Radio and Fixed Access Network functions (5G/6G, FTTH), Core Network functions (Control and User Planes) or Network Controllers (Transport SDN, RAN RIC), as well as Telecom Services platforms.

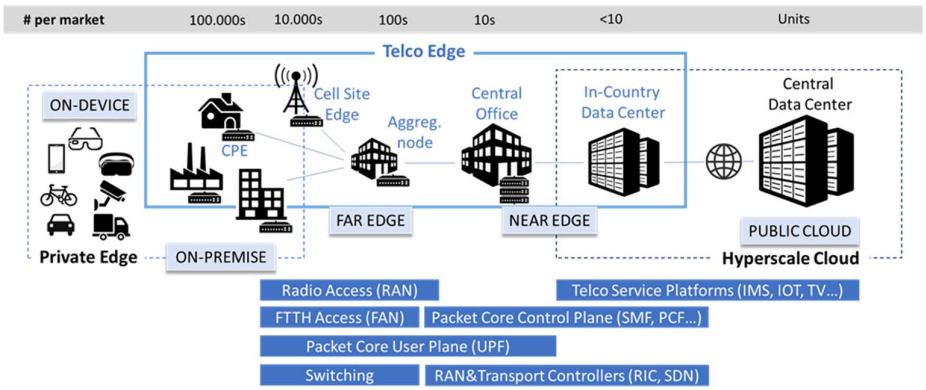


Figure 3: Telcos need hundreds of edge nodes to support the evolution of the network towards the cloud.

In this context, the cloud-native infrastructure offers a great opportunity to make network *deployment simple, dynamic, and efficient*, enabling a consistent computing platform that facilitates sharing of hardware and connectivity resources among different applications. To fully benefit from the opportunities that cloud native presents, the telco cloud layer should align with well-established cloud native technologies and operating processes (such as current architectures based on automated multi-cluster Kubernetes deployments on bare metal) and be managed with a GitOps approach, including mechanisms like Infrastructure as Code.

Moreover, the telco cloud, due to its support for critical national infrastructures such as telecommunication networks, needs to be designed as a *sovereign cloud* in order to address the goals defined by the EU (as referred to in Section 2 of [2]):

- Cybersecurity: critical infrastructures and essential services require higher levels of security and an increased operational capacity to prevent and manage attack scenarios. Security enhancement, implementing features like Internal ciphering or multi-tenancy, requires the adoption of certification schemes like the ones defined by the European Union Agency for Cybersecurity (ENISA).
- Trustworthy data processing: the full control of features is a first step to guarantee interoperability, data protection, portability, and trusted data sharing among companies.
- Strategic autonomy: supported by a diverse supply chain. Virtualisation/containerisation solutions need to be hardware-agnostic (i.e. run on hardware from different providers) to avoid vendor lock-in and provide interoperability and portability, again to avoid lock-in.
- Digital sovereignty: avoiding dependency on foreign entities and regulations for digital technologies and services at different levels: supply, operation, data and supporting infrastructure, including hardware, software, and communications.

Some of these goals are common to other types of use cases, such as the sovereignty requirements that are applicable to any other critical infrastructure or service and can be achieved using existing open source standards, which enable a high degree of *interoperability* and allow switching and *portability* in a distributed environment. Nevertheless, the nature of the network functions and services demands specific *technology requirements* from the distributed cloud-native architecture, as explained below.

Technology requirements

R1. Enhanced performance to guarantee that services are allocated with direct access to the hardware to speed up data transfer and meet the latency required in the infrastructure. This includes supporting Enhanced Placement Awareness (EPA) capabilities such as Multus-DPDK[2], SR-

[2] DPDK: Data Plane Development Kit. It provides libraries and network controllers to accelerate packet processing in the user plane.

IOV[3], CPU-pinning[4], and Huge pages[5], although this may change with the evolution of microelectronics.

R2. Support for PTP [6] to cope with the specific synchronisation protocol defined in O-RAN.

R3. Hardware acceleration such as SmartNICs, GPUs, FPGAs, etc. to enable the cost- and power-efficient implementation of demanding network functions like RAN L1 or edge applications.

R4. Bare metal support to facilitate lightweight container clusters that run directly over the hardware, necessary due to the power and space restrictions of some telco edge locations.

R5. Virtual Machine support (e.g. support for kubevirt) to cope with hybrid use cases requiring traditional virtualisation and containerisation together. In general, telecom networks require long-term support for the different infrastructure components, including the cloud.

R6. Specific form-factor support to enable virtualisation using distinct hardware that may be required for specific edge services (e.g. lightweight servers intended for microsites requiring only frontal access).

The telco cloud, as a platform that can host edge services in a multi-site distributed environment, also faces *operational challenges*. The *infrastructure lifecycle management* of hundreds of edge-cloud sites raises potential issues that must be tackled efficiently to ensure that the overall strategy is sustainable from an economic perspective. These requirements are listed below.

Infrastructure lifecycle management requirements

R7. Bare metal automation provisioning (e.g. via CAPI) for dynamic deployment of a massive highly distributed computing infrastructure.

R8. Multi-clustering to enable monitoring of this distributed infrastructure.

R9. Networking automation (e.g. via SONIC) for dynamic deployment of the underlying computing and networking infrastructure. It incorporates *advanced granular traffic steering* to manage the lifecycle of the computing infrastructure and the network functions deployed on top, and to automate large-scale application deployments (including network services) without service impact, which is a key need given the high number of edge nodes and applications.

R10. Automation and control from remote locations and limited on-site manual intervention should apply to all lifecycle phases, especially relevant for telcos due to the scale and capillarity of their deployments. It should apply to instantiation and configuration of a new site, and site scaling (e.g.

[3] SR-IOV: Single Root Input/Output Virtualisation. It allows the sharing of an Input/Output resource (PCI) among several virtual machines.

[4] CPU pinning, also known as CPU affinity, is a technique that allows a process or a thread to be bound to a specific CPU or core.

[5] Hugepages is a Linux kernel feature that allows the use of bigger memory pages to improve system performance.

[6] Precision Time Protocol (PTP) is a protocol used to synchronise clocks throughout a computer network.

adding new servers using factory-mounted pre-configured server racks, well established for big data centres and even more relevant in a highly distributed computing infrastructure in which physical activity has a much higher cost), configuration updates, software upgrades (OS images, Kubernetes versions, etc.).

R11. Specific infrastructure customisations that are required by certain telco or customer application types, like the hardware accelerated nodes used for baseband processing. Their proliferation should be nevertheless avoided, encouraging their consolidation in a limited set of immutable blueprints/configurations to avoid an unmanageable environment that may waste the benefits of cloud technologies.

R12. Multi-cloud orchestration since telecom operators need to distribute their network functions across a multi-provider, multi-technology computing environment, following a hybrid multi-cloud approach. To this end, they must be able to manage container clusters (or whatever other virtualisation technologies arise) in different public and private clouds, over several cloud technologies.

These requirements, while valid in general, are especially important for the telco industry which needs to incorporate and manage a plurality of network functions of diverse nature that are still in the process of becoming truly cloud native. Additionally, this kind of cloud presents requirements in relation to the *application lifecycle management*.

Application lifecycle management requirements

R13. Dynamic application orchestration to determine the right location for a given service, to configure the selected site with the right capabilities for the service (e.g. GPU support for video-related services, or for radio base band processing) and to deploy the software image in the selected site once configured.

R14. Edge discovery and *mobility management* to support the user device in the identification of the optimum edge node to serve a certain application based on its location and to re-arrange the connectivity configuration. These contribute to an enhanced edge experience and its recalculation as the users move across different locations. The latter requires a bonded and seamless interaction with the network.

R15. Federation to allow multiple providers to collaborate to deliver edge computing services of public interest (e.g. V2X, assisted, or autonomous driving) across different operator markets and geographies.

R16. Portability, enabling operators to port network functions from one cloud to another without additional (or with very limited) adaptation, integration, or testing effort. This requirement emerges from the need to provide service and business continuity in an essential public service

(telecommunications). In case a cloud technology or provider fails, there needs to be a mechanism to port all the workloads on that technology to another in the shortest time.

2.2 Edge Application Use-Case

In addition to enhancing internal capabilities, telecom operators have a unique opportunity to leverage cloud and edge computing capacities for customer[7] utilization. This offers a new potential revenue stream and, at the same time, meets the growing demand for localised, high-performance computing resources.

In recent years, the need to increase the proximity between cloud services and end-users has emerged with growing clarity. Services are moving from remote data centres to be deployed in multiple locations in a market, providing the benefits of large-scale computation capabilities and the readiness, immediacy, and privacy of smaller computing nodes close to the edge application users.

As described above, telcos will need to deploy *hundreds of edge cloud nodes* that can host third-party applications, like over-the-top connectivity (e.g. firewalls, SD-WAN, etc), virtualised private networks, content delivery networks, video analysis for security cameras, control platforms for robotised production plants, autonomous or assisted driving, drones or AGVs, AI training and inferencing solutions, AR/XR gaming, sensor networks, or metaverse service platforms.

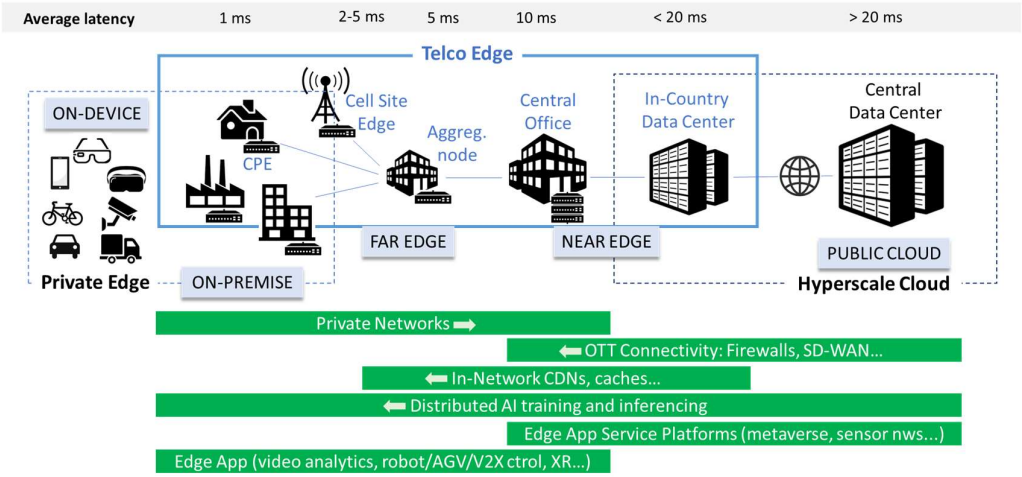


Figure 4: Edge Applications can be hosted in different Telco Cloud locations providing scale benefits.

The delivery of cloud and edge services to customers (external applications) on the telco cloud infrastructure brings its own set of challenges, as introduced in the following paragraphs.

[7] A *customer* is a consumer of the edge computing capacity, that is, an edge application developer or an ISV that deploys its software at the edge.

Edge application use case requirements

R17. Multi-tenancy to enable sharing of the infrastructure and its application services by multiple entities.

R18. Isolation towards enhanced security to protect the applications and virtually allocated edge resources of each customer from the interaction of other customers using the same edge infrastructure. This is especially important when the service is used for network functions, as they support a critical national service.

R19. Universality (as a standard and universal interface, an "as-a-Service"-type API), to increase attractiveness for the customers so that they can develop an edge application and deploy it in different edge nodes from multiple operators worldwide without any adaptation. This universal API will tackle different (i.e. multiple) providers in terms of deployment and management (tackling diverse cases such as containers, serverless, etc).

R20. Federation to facilitate application lifecycle management for customers, allowing them to deploy and configure applications over a global multi-provider footprint by contacting a single provider.

R21. Data security and regulatory compliance, through a uniform trust framework to give customers the confidence to deploy applications without increasing regulation and business compliance hurdles.

R22. Dynamic resource allocation management to facilitate "consume-as-you-go" models comparable to those available in public clouds. This enables edge applications to automatically scale up and down their resource allocations according to actual usage and efficiently use the limited computing resources at the edge.

R23. Transparency and reliability as key factors to maintain customer trust and ensure long-term engagement.

It should be noted that, even though this section covers the requirements for the infrastructure hosting telco services, some of these requirements may apply to other industries that demand a distributed computing architecture such as RAN-oriented hardware acceleration that could be used for AI-based applications, networking automation that applies to large scale in-vehicle application deployments, or specific form factors that could apply to edge use cases in constrained environments like agriculture.

SECTION 2: INITIATIVES ADDRESSING TELCO REQUIREMENTS

2.1 OpenNebula

OpenNebula is a European open source cloud and edge computing solution licensed under Apache 2.0 [5] with regular active contributions from the OpenNebula User Community [6]. It has developed reference architectures for an Open Cloud [7] and an Edge Cloud [8], and provides an abstraction layer on both the private cloud/edge infrastructures and those of external providers. The platform delivers native services like distributed edge-cloud computing and bare metal provisioning (tackling *R4, R7*). Furthermore, it implements edge computing features including *R17* (multi-tenancy) and *R18* (isolation) as native functionalities, while complying with *R1* (enhanced performance support), *R2* (support for PTP), *R3* (hardware acceleration support), *R5* (virtual machine support). The *R20* (federation) requirement will be delivered by OpenNebula as part of the recently awarded IPCEI-CIS [23] and ONEedge5G [9] projects.

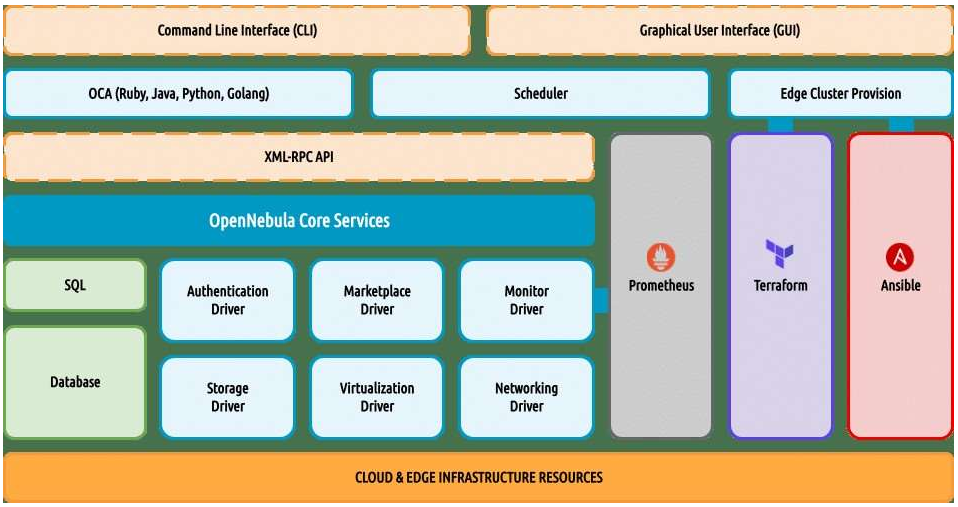


Figure 5: OpenNebula overview.

OpenNebula proposes two main architectures to modernise telco's existing networks, simplify network operations, and speed-up deployment:

- *Highly distributed NFV deployment,* which tackles *R12* (multi-cloud orchestration). It is realised through a single cloud front-end to manage tens to hundreds of geo-distributed clusters (PoPs with edge nodes) with minimal hardware infrastructure, local storage, and DPDK or SR-IOV technologies for high-performance throughput.

- *5G edge deployment*, which tackles *R10* (automation and control). It is facilitated through the installation of micro- data centres in edge nodes at 5G locations in order to host O-RAN and offer edge cloud services following a Multi-access Edge Computing (MEC) architecture. It allows operators to deploy third-party edge services easily and seamlessly, ensuring the required network isolation and security while providing the application owner full control of their 5G/edge service deployment.

2.2 ANUKET

Anuket (Linux Foundation) [10] delivers a common model, standardised reference infrastructure specifications, and conformance and performance frameworks for virtualised and cloud native network functions, enabling faster, more robust on-boarding into production, reducing costs and accelerating communications digital transformations. Anuket addresses a wide range of use cases from core to the edge. Anuket artefacts include an integrated, tested, and validated open software reference infrastructure used to design a conformance framework and verification program.

R19 (universality) is covered by the "Reference Architecture" deliverables of Anuket (RA-1 and RA-2). They provide a common, universal language and a set of platform architectures to be leveraged across the entire telecom industry. *R23* (transparency / reliability) is tackled by the testing and conformance framework, a suite of testing tools that addresses both functional conformance and performance issues. These tools are freely available for operators and vendors.

Anuket is part of LF Networking, which was formed in collaboration with GSMA (Cloud iNfrastructure Telco Taskforce, CNTT) and partners with other standards bodies, open source communities including CNCF, ONAP, and OpenStack, and industry-leading network operators and NFVI/VNF suppliers with a global scope. Anuket Reference Architecture and Conformance test suites are implemented in Sylva project, which is further detailed below.

2.3 SYLVA

The main carriers in Europe and leading network function providers launched the Sylva project [11] at Linux Foundation Europe in November 2022, addressing telco and edge requirements and use cases, with the intention of increasing telco cloud convergence and operational efficiency to benefit all industry players.

Even though the use of cloud native technologies promotes standardisation, the traditional deployment model applied by the Telco industry leads to fragmentation for the deployment of applications. This is caused by the different methods each platform and each network function vendor has used to meet the technology and infrastructure lifecycle management requirements

described above, as well as the use of proprietary k8s extensions and even specific physical compute as shown in the following picture.

Today, this fragmentation forces vendors to develop their Network Functions (NFs) to be able to certify against multiple cloud layers and as a result, operators deal with multiple infrastructures, which increases operational complexity.

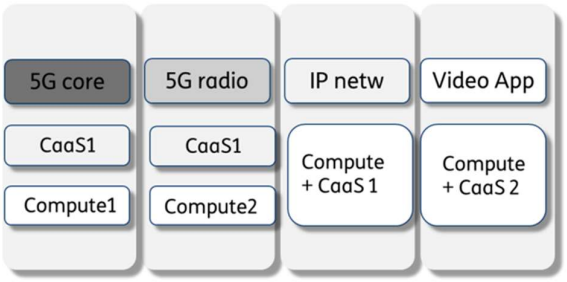


Figure 6: Fragmentation in the deployment of Telco workloads.

The edge cloud needs to evolve at the same speed as the internet and the public cloud, while continuous upgrades and evolution of software are required, for instance new features or security patches. CaaS software and application software will need to be updated regularly. The hyperscalers can provide such capabilities to both operators and vendors, but only as proprietary solutions (no interoperability, strong lock-in), compromising the goals of *strategic autonomy* and *digital sovereignty* described in section 2.

The Sylva project delivers a *cloud software framework*, addressing the specific technical challenges of the telco infrastructure layer, integrating these solutions with existing open source components to make them "production grade" [8], and develops a *reference implementation* of this framework. Moreover, an *integration and certification program* provides the means to validate cloud implementations based on it and the commercial network functions against it.

As an open source project, it leverages existing cloud native community-driven open source projects, reference implementations, and conformance test frameworks such as Kubernetes, Anuket, or Nephio, and its roadmap is driven by a European Telecom ecosystem, which is open and applicable worldwide.

This project delivers CaaS capabilities to address use cases like 5G Core, Open RAN, and Edge, and will be the basis for a common infrastructure among European operators. This common infrastructure will facilitate federation and integration of edge applications, decrease integration and testing time and costs to both vendors and operators, and is expected to foster digital innovation.

Sylva delivers some extended capabilities required by telco use cases covering technology requirements *R1 - R6* (enhanced performance, synchronisation, acceleration, bare metal support, multi-clustering, efficiency, form-factor), infrastructure lifecycle management (*R7 - R9*) and multi-tenancy and isolation requirements (*R17, R18*). More details can be found in the "Sylva Technical Principles" architecture [12].

[8] Production-grade means that it can be used by operators, network vendors, and cloud providers to create commercial products.

2.4 NEPHIO

Nephio (Linux Foundation) [13] provides a Kubernetes-based intent-driven *automation of network functions and the underlying infrastructure* that supports those functions. It allows users to express high-level intents and provides intelligent, declarative automation that can set up the cloud and edge infrastructure, render initial configurations for the network functions, and then deliver those configurations to the right clusters to get the network up and running.

A distributed cloud enables on-demand, API-driven access to the edge and demands a new approach to handle the complexity of provisioning and managing a multi-vendor, multi-site deployment of interconnected network functions across this on-demand distributed cloud.

The solution is intended to address the initial provisioning of the network functions and the underlying cloud infrastructure, and provide Kubernetes-enabled reconciliation to ensure the network stays up through failures, scaling events, and changes to the distributed cloud. Nephio breaks down the larger problem into two primary areas:

- Kubernetes is used as a *uniform automation control plane* in each site to configure all aspects of the distributed cloud and network functions at each layer of the stack. Nephio is establishing open, extensible Kubernetes Custom Resource Definition (CRD) models for each of the three layers of the stack: (i) cloud infrastructure resource automation, (ii) workload resource automation, and (iii) workload configuration. It provides tools and libraries to assist vendors with integrating existing Yang and other industry models with Nephio, in conformity with standard interfaces like O-RAN 01 or 3GPP. Providing the same tooling at every layer, it enables the automation of interrelated configuration between those layers.
- This framework leverages Kubernetes *declarative, actively reconciled methodology* along with machine-manipulable configuration to minimise the complexity of these configurations. It implements the Configuration-as-Data approach in the scope of configuration management. This enables users to author, review, and publish configuration packages which may then be cloned and customised to deploy network functions. This customisation can be fully automated, or mix-and-match automated and human-initiated changes without conflicts and without losing the ability to easily upgrade to new versions of the packages.

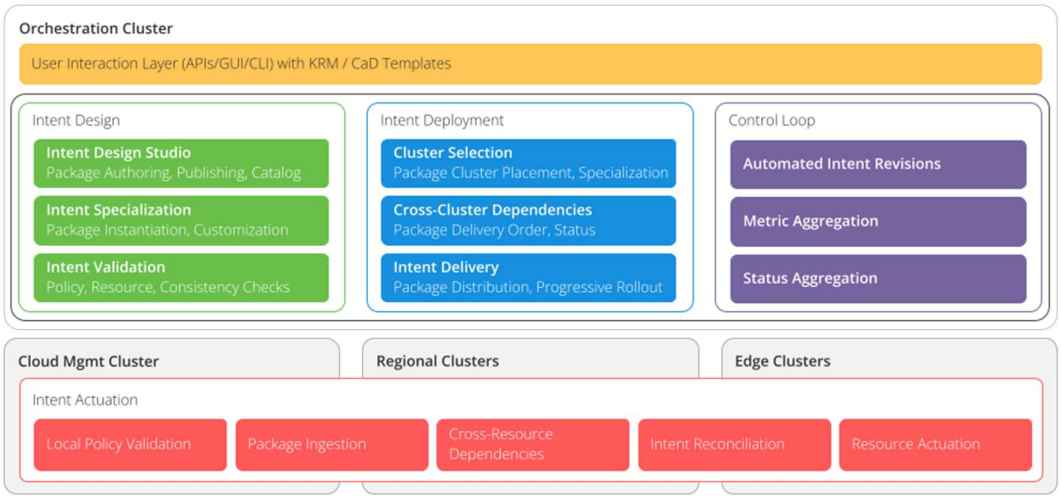


Figure 7: Nephio's Reference architecture.

Nephio simplifies the overall automation and enables declarative management with active reconciliation for the entire stack, addressing requirements like the automation of configuration updates and software upgrades (*R10*), specific infrastructure customisation (*R11*), and multi-cloud orchestration (*R12*).

The CRDs and operators used for cloud infrastructure resource automation can exploit existing Kubernetes-based ecosystem projects as pluggable southbound interfaces (e.g. Google Config Connector, AWS Controllers for Kubernetes, Azure Service Operator, and via Cluster API for Sylva), providing an open integration point and more uniform automation across those providers.

Today, workload resource management effectively requires complex Infrastructure-as-Code templates that are purpose built for specific network functions. Taking a Configuration-as-Data, Kubernetes CRD approach, capturing configuration with well-structured schemas, allows the development of robust standards-based automation.

2.5 CAMARA

The CAMARA project (Linux Foundation/GSMA) [14] aims to define, develop, and test *standardised network APIs to expose the telco network capabilities* for customers in a seamless manner. It addresses the universality requirement (*R19*). Through standardised APIs, the CAMARA project facilitates the implementation of applications for developers, utilising the telco network regardless of the context.

Telco capabilities in 4G and 5G systems, and in related infrastructures (e.g. edge computing, distributed AI) and platforms (e.g. billing, identity management, etc.) are exposed for use by external systems. These functions enable the collection of network and systems information and its

utilisation for their configuration. The on-demand, secure and controlled exposure of these capabilities paves the way for transforming operator networks into service enablement platforms, facilitating the application-to-network integration, which will be key to deliver enhanced and service-tailored customer experience in the 5G era. CAMARA is an open source project within Linux Foundation to define, develop, and test the APIs. CAMARA collaborates with the GSMA Operator Platform Group that aligns API requirements and publishes API definitions and APIs, and with TMForum that provides the complementary Operate APIs that facilitate the connection of the operators' API gateways with the different sales channels (marketplaces, developer environments, aggregators, etc.).

The CAMARA project has a large backlog of APIs. The technical contributions are organised around each type of API. All those APIs contribute to covering the requirements *R16, R19, R21* and *R23*, as the descriptions of the APIs are public, providing a simple and standard user experience that facilitates portability and a trusted secure environment. The APIs and procedures are tested and documented, meeting the transparency and reliability requirements.

2.6 OPI

Open Programmable Infrastructure (OPI) [15] is a community-driven, standards-based open ecosystem for next-generation architectures and frameworks based on DPU (Data Processing Unit) / IPU (Infrastructure Processing Unit) technologies. This includes collaborative development in documentation, testing, integration, and the creation of artefacts that aid the development, deployment, operation, or adoption of the OPI project.

OPI establishes open standards for DPU/IPU-like technologies, creating a vendor agnostic framework and architecture for DPU/IPU-based software stacks, facilitating the reuse of existing APIs or defining a set of common APIs when required for DPU/IPU hardware, and providing implementation examples to validate the architecture/APIs.

Lifecycle management of DPU/IPU hardware is one area of contributions, focused on common APIs around the DPU/IPU hardware, DPU/IPU hosted applications, provisioning software, orchestration software, storage use cases and networking use cases. Specifically, the networking contributions are focused on the networking offload to a DPU/IPU stack of several use cases such as cloud, router (e.g. EVPN Gateway), security/encryption (e.g. IPSEC) and other networking functions.

OPI brings together all the major hardware vendors, applies open source development principles, and is open to new members and contributors. It defines a vendor-agnostic integration framework that includes software abstraction layers and frameworks as well as vendor agnostic services, lifecycle management, and observability, facilitating hardware acceleration (*R3*), automation and control (*R10*), portability (*R16*), and dynamic resource allocation (*R22*).

2.7 StarlingX

StarlingX [16] (Open Infra Foundation) offers a fully *integrated, highly reliable edge infrastructure software stack* that enables critical infrastructure services for network performance, low latency, and high bandwidth in edge or IoT environments. It integrates Ceph for storage, OpenStack for cloud infrastructure, Kubernetes for container orchestration, and supports technologies like DPDK and SR-IOV for performance enhancements.

The platform directly contributes to the telco cloud by providing:

- Enhanced performance and bare metal support, catering to low-latency, high-bandwidth applications.
- Integration of PTP and hardware acceleration technologies.
- Support for virtual machines and containers, aiding in the transition to cloud-native network functions.
- Automated lifecycle management of edge sites, including bare metal provisioning and multi-cloud orchestration.
- Solutions for dynamic resource allocation, application orchestration, and security, addressing multi-cloud and federation challenges.

StarlingX addresses several key requirements for the telco cloud, focusing on technology *R1 - R6* (enhanced performance, synchronisation, hardware acceleration, bare metal and virtual machine provisioning, and specific form factors), and challenges in edge infrastructure management (*R7 - R10*). It also contributes to multi-cloud orchestration (*R12*), dynamic application orchestration (*R13*), portability across clouds (*R16*), and multi-tenancy and isolation challenges (*R17, R18*).

StarlingX operates in a rich ecosystem involving the OpenInfra Foundation and the larger open source community, with contributions from significant industry players like Intel and Wind River. Its development is heavily influenced by the needs of telecommunications and industrial IoT sectors.

2.8 ETSI-NFV

Founded in November 2012 by world's leading telecom network operators, ETSI-NFV [17] became the home of Network Functions Virtualisation (NFV), developing the required *standards for NFV transformation*, incorporating the latest technologies as well as sharing their experiences of NFV implementation and testing in multi-vendor environments. From Release 4 [18], additional support for containerised VNFs towards more cloud-native deployment and automation has been specified. Release 5 [19] addresses how the NFV framework can support virtualised Radio Access Network (vRAN) use cases, covering key issues like acceleration abstractions, hardware acceleration resource management, and transport network management, including time synchronisation.

ETSI-NFV is currently considering how NFV will evolve [20]. To this end, new requirements for the telco cloud are considered: (i) unified network management, (ii) use of latest cloud-native, IT, automation, and AI open source software, and (iii) multi-vendor interoperability and migration among different clouds.

NFV envisages the implementation of Network Functions (NFs) as software-only entities that run over the NFV infrastructure and identifies three main working domains: (i) *Virtualised Network Function*, network function software that can run over the NFVI, (ii) *NFV Infrastructure* (NFVI), supporting the execution of the VNFs, including the physical resources and their abstraction layer, and (iii) *NFV Management and Orchestration*, enabling orchestration and lifecycle management of resources that support the infrastructure virtualisation and the lifecycle management of VNFs.

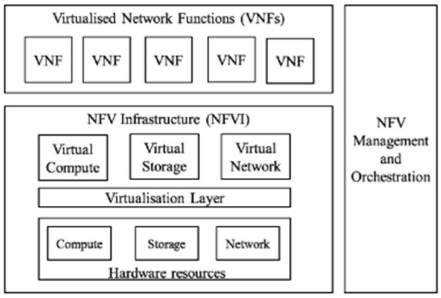


Figure 8: Main ETSI-NFV domains.

A Network Service (NS) is a composition of NFs arranged as a set of functions and/or NSs. The NFs can be physical (PNF, software workload tightly coupled to the hardware it is deployed on), virtual (VNF, deployed on a virtualised machine, VM) or containerised (CNF, deployed on a container cluster, realised as VMs or bare metal). The main elements in the architecture are the following:

- *Virtualised Infrastructure Management* (VIM) that controls and manages the interaction of a VNF with computing, storage, and network resources, as well as their virtualisation.
- *VNF Manager* (VNFM), being responsible for VNF lifecycle management (e.g. instantiation, update, query, scaling, etc.).
- *NFV Orchestrator* (NFVO) that orchestrates and manages the NFV infrastructure and software resources and realises the network services on NFVI.

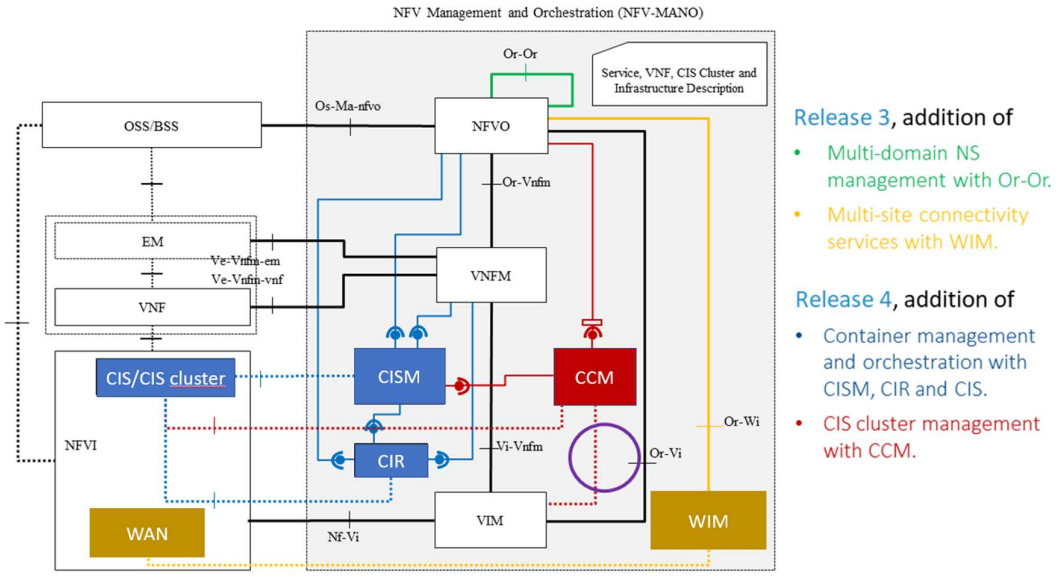


Figure 9: ETSI-NFV Architecture, main interfaces, and technical specifications (Release 4).

As depicted in the figure above, Release 3 adds relevant functionality to manage Network Services across multiple domains (Or-Or interface) and connect services and NFs across multiple sites (WIM, WAN). Release 4 is extended to manage containerised VNFs with three new elements: (i) *Container Infrastructure Service Manager* (CISM) to manage containers (e.g. Kubernetes control plane), (ii) *Container Image Repository* (CIR) to enable storing of container images (like the Docker Registry), and (iii) *Container Cluster Manager* (CCM) to handle container clusters over different environments (such as Kubespray, kubeadm, and cluster-api).

ETSI-NFV specifications already cover technology requirements including bare metal support (*R4*) and virtual machine support (*R5*), as well as infrastructure lifecycle management requirements such as multi-clustering support (*R8*), support for networking automation (*R9*), automation and control (*R10*), and multi-cloud orchestration (*R12*). Release 5 is expected to address additional requirements such as support for accelerator abstraction (*R3*) and transport management, including time synchronisation (*R2*). In its new phase it facilitates unified network management and automation (*R10*), and multivendor interoperability and migration among different clouds (*R13, R16*).

To realise NFV in commercial products, ETSI NFV joined efforts with open source communities (OPNFV and Anuket) to release NFV integration solutions based on existing IT and open source software. Given the importance of network and service orchestration for service providers, other open source communities such as OSM, OPEN-O, and ONAP have built upon the concepts from NFV.

Releases 4 and 5 required close coordination with CNCF (Kubernetes, Helm...) and O-RAN (acceleration).

2.9 ETSI-OSM

Open Source MANO (OSM) [21] is the first open source effort hosted by ETSI focused on the development of an *open implementation of the NFV MANO* (Management and Orchestration) stack aligned with ETSI-NFV. It is used as the reference implementation and base for commercial distribution of telco cloud orchestration platforms, and for research and demonstration of new orchestration capabilities.

OSM provides a production-quality Open Source NFV MANO that considers the full lifecycle of the network functions (day-1 and day-2) as well as related infrastructure components such as Kubernetes clusters and cloud-provided PaaS services, in a *cloud- and technology-agnostic* manner. It models the deployment characteristics and the associated lifecycle of network services and network functions through both *service and resource orchestration*.

OSM stack delivers cloud-based network service lifecycle management capable of exploiting openly published information models, available to everyone, suitable for all types of network functions, operationally significant and independent of the cloud infrastructure used. OSM is aligned and provides experience-based feedback to ETSI NFV.

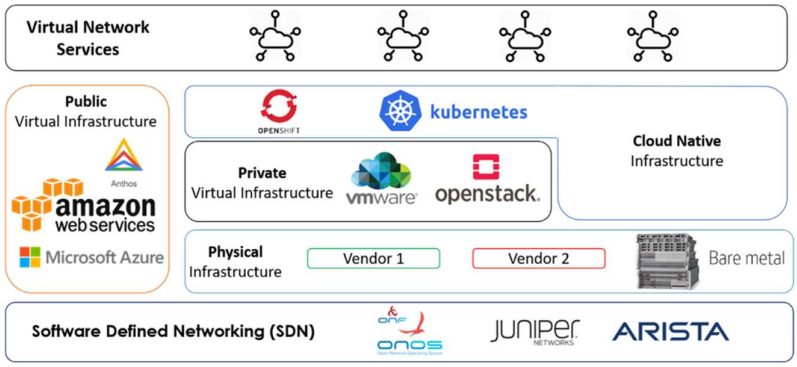


Figure 10: OSM orchestration of a fully hybrid infrastructure.

OSM enables the creation of networks on demand ("Network-as-a-Service" or NaaS) for direct exploitation by the service provider or potential commercialisation to third parties. In that sense, OSM works as a Network Service Orchestrator allowing, besides the creation of network services on demand, the control of the lifecycle and operations of the network service via calls to OSM's northbound API and the monitoring of its global state. OSM consumes services provided by the platform(s) in charge of the cloud infrastructure and the platform(s) in charge of the Software-Defined Network (for different inter-DC connections), and once assembled, configures and monitors the constituent network functions to control the lifecycle of the network service to be offered on demand.

OSM exposes its orchestration services via a northbound interface based on ETSI NFV specifications. In this case, it consumes openly published data models and translates those requirements into specific calls to the target clouds and environments, providing a cloud-agnostic modelling that grants portability across clouds. In addition, OSM monitors the deployed resources and triggers corrective actions to deviations from expected behaviour, or scales as needed.

OSM delivers a regularly updated MANO stack reference implementation (two releases per year) that implements a network virtualisation orchestration open source solution for the deployment of workload images onto virtualisation environments, contributing to requirements (*R13* and *R22*).

Since its conception, OSM has addressed issues related to platform awareness (*R1*) and acceleration technologies (*R3*), and supports any mix of deployment styles, combining bare metal (*R4* and *R7*) and virtual machine support (*R5*), as well as the seamless integration of physical network functions (PNFs) and fine-grained control of function placement rules (*R13*). Apart from the automation and monitoring capacities incorporated in the OSM platform, it provides a powerful plug-in framework to integrate other network and cloud orchestration tools to facilitate automation and dynamic monitoring (*R7, R8, R9, R10* and *R11*). OSM has probably the most complete set of multi-cloud orchestration capacities among telco cloud orchestrators, including dynamic orchestration features (*R12, R13*). The use of standards for modelling and lifecycle management of network functions and the use of standard interfaces promotes portability (*R16*).

Development in OSM is based on accepted open source working procedures [22] and hosted by ETSI. The development community is open, made of ETSI and non-ETSI members, with 44 members (including 7 operators) and 109 participants, and 48 companies who have contributed to the codebase. To date, code for OSM Releases has been downloaded over 60,000 times from more than 85 countries and at least 39 EU-funded 5G/6G research projects are using and contributing code and feedback to OSM.

2.10 ETSI-ZSM

ETSI Zero-touch Network & Service Management (ZSM) enables largely autonomous networks, suitable to be driven by high-level policies and rules, and capable of self-configuration, self-monitoring, self-healing, and self-optimization without human intervention. It has defined a comprehensive architectural framework for end-to-end automation, comprising an arbitrary number of domains (*R9, R10*) and identifying the service classes and essential functions for automation integration and loose coupling of management services (*R11*). ETSI ZSM addresses aspects related to the application of closed-loop and intent-based principles to network automation, as well as the lifecycle management of these technologies (*R13*) in multi-domain environments. The architecture framework for multi-cloud environments is currently being explored (*R12, R16*).

ETSI ZSM aims to provide a holistic end-to-end network and service management concept, building a flexible service-based network and service management framework that supports cross-domain end-to-end management and provides enablers for closed loop automation and for data-driven management algorithms that can be based on ML/AI. Additionally, the ZSM framework reference architecture defines a set of building blocks that collectively enable the construction of more complex management services and management functions using a consistent set of

composition and interoperation patterns. Management domains provide the means to separate management concerns, considering boundaries of different natures (technological, administrative, organisational, geographical, etc.). Every management domain provides a set of ZSM management services, realised by management functions that expose and/or consume a set of service endpoints. An end-to-end service management domain is a special management domain responsible for the cross-domain management and coordination.

The group is committed to considering a series of technical matters connected to the telco cloud goals, especially regarding the support for cloud-network integration, with a special focus on edge scenarios, the Network-as-a-Service (NaaS) paradigm and the APIs for network and service capability exposure, and the consideration of extended management capabilities, such as Network Digital Twins and Generative Models.

ZSM comprises 49 ETSI members (including 14 network operators) and 28 participants (non-ETSI members).

2.11 IPCEI-CIS

Europe has emphasised the need to strategically invest in the next generation of cloud and edge capacities on several occasions, such as in the European Strategy for Data [24] and in the Member States' Joint Declaration on Building the Next-Generation Cloud in Europe [25]. In this context, 12 EU Member States collaborate towards a coordinated industrial response of significant importance to Europe: the Important Project of Common European Interest on Next-Generation Cloud Infrastructure and Services (IPCEI-CIS) [23]. This ambitious integrated project will first industrially deploy a fundamentally new and innovative data processing production infrastructure based on breakthrough technological advancements and the development of advanced industrial products and services, while contributing to existing European initiatives and policies, in particular to the European Green Deal [26], the European Industrial Strategy [27], the European Data Strategy, and the Digital Compass [28].

The IPCEI-CIS integrated project will be realised as a set of individual projects with more than 100 companies participating from 12 EU member states, aiming to deliver an advanced technical solution for a multi-provider cloud-edge continuum, based on open source, standard, secure and green technologies, that covers several telco cloud requirements.

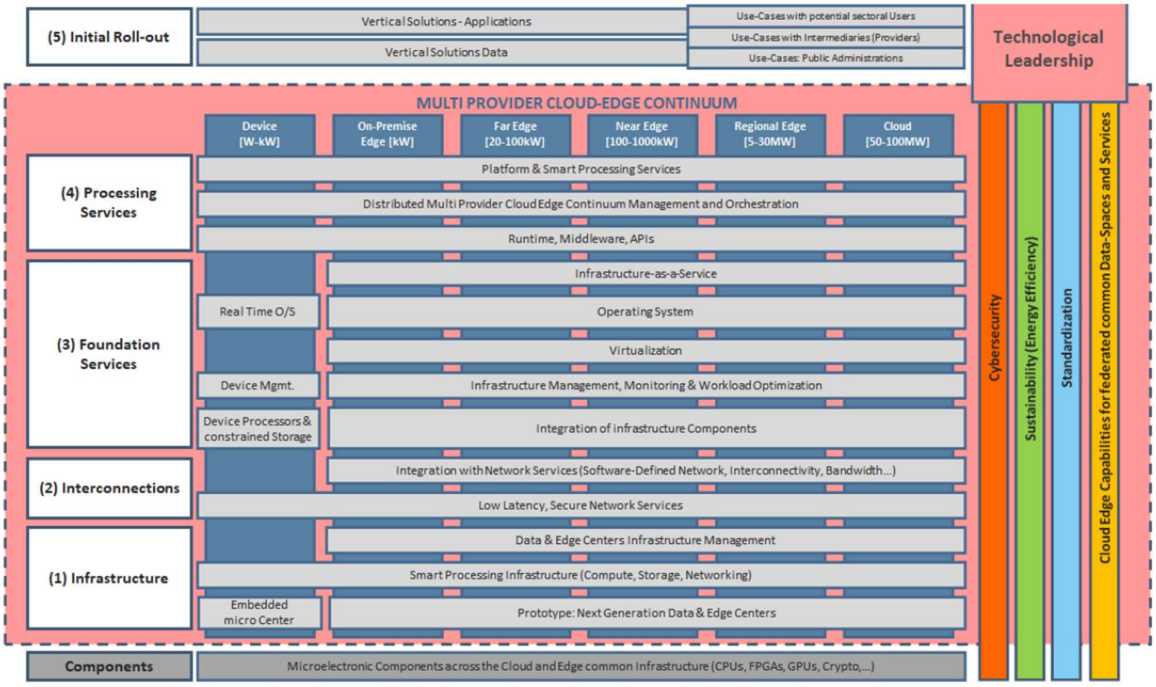


Figure 11: Value chain for the IPCEI-CIS [23].

The integrated project considers all types of computing nodes (on-premises, far, near edge and regional edge nodes and cloud nodes). Several IPCEI individual project portfolios include contributions to some of the initiatives described in this section, such as Sylva, CAMARA, and OpenNebula, and encompass the deployment in production environments of uses cases on top of a Sylva-based telco cloud using CAMARA APIs to get access to edge connectivity and computing services.

The IPCEI-CIS contributes to automation and orchestration requirements (*R10, R12, R13*), provides infrastructure blueprints for near and far edge (*R11*), delivers the mechanisms for a multi-provider edge cloud continuum (*R15, R16, R19, R20, R23*), designs and develops an optimal integration with the network (*R14*) and has a special focus on security (*R21*) and sustainability (*R22*). Furthermore, it indirectly covers several technology and infrastructure lifecycle management requirements (*R1- R9, R17, R18*) by supporting part of the development at Sylva project.

2.12 Vates VMS

Vates VMS (Virtualisation Management Stack) is a European, open source full virtualisation stack. It provides a full server virtualisation stack including an actively and independently developed hypervisor, XCP-ng [29], that is hardened Xen-based, and a management, orchestration and backup plane that is extensible and agentless [30]. As such, it fully covers the *R5* features (virtual machine support) as it is its own hypervisor but also supports another hypervisor such as

XenServer. Through specific investments it is capable of addressing *R3* and *R4* through GPU support, and has the ability to run on DPU [31] and the capacity to handle fine-grained backup and infrastructure resilience.

Vates VMS also provides the ability to automate entire aspects of the infrastructure management through task automation, ACLs, and the enablement of multi-cloud initialisations (*R10, R11, R17*). Through its advanced features providing fine-grained backup and replication across the network as well as logic isolation of virtual environments, the stack meets the requirements *R18, R21, R22, R23.*

Vates VMS is successfully used in edge computing use cases, most notably in the fields of energy, manufacturing and distributed infrastructures for public services. In that regard, Vates is currently working with a consortium on the ReNESENS research project (French BPI-funded initiative) that aims at designing edge computing specific appliances with advanced backup technologies.

Vates VMS is also actively porting the Xen core technology and its VMS stack on the Arm and RISC-V processor architectures, to deliver the best possible integration of hardware and software for clouds and data centres and thus achieve the hybrid cloud-edge continuum of infrastructure, services, and data.

2.13 Other related activities

Linux Foundation Networking have recently created the *Cloud Native Telco Initiative* [32] with the intention of defining the procedures, tools, and certification schemes to verify that CNFs are actually cloud and Kubernetes native. It builds on the work of LFN's Anuket and CNCF's CNF Test Suite and certification programs, aiming to forge a consensus on cloud native operating models for telecom.

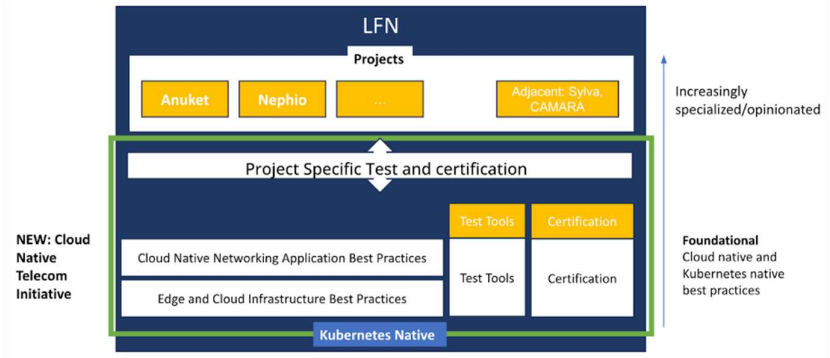


Figure 12: CNTI building blocks to test and certify CNFs.

CNTI, supported on other initiatives as the ones mentioned above, contributes solutions such as Anuket, Nephio, Sylva or CAMARA, cloud-native networking and automation and control (**R9**, **R10**), while with respect to true Cloud- and Kubernetes-native CNFs, it facilitates universal and portable services (**R16**, **R19**) and brings transparency and reliability (**R23**).

SECTION 3: CHALLENGES AND REQUIREMENTS TO BE ADDRESSED

The initiatives described in the previous section focus on standardising a telco cloud environment that can host both network functions and innovative edge services based on the use of standard interfaces and, optionally, open source solutions. However, there are notable gaps in current efforts that require attention and these are summarised in the following paragraphs.

C1. Multi-cloud Orchestration

One of the main aspects to consider in the edge cloud architecture is the fact that the underlying infrastructure where the edge applications will be running must be a cloud-native infrastructure, composed of a set of environments distributed in different locations and potentially served by several technology providers. Managing different cloud environments (private and public) with different cloud technologies is still an unsolved issue.

ETSI-NFV has taken some steps in this direction with the definition of the CCM (Container Cluster Manager). Moreover, ETSI-NFV has produced several commercial implementations, but does not address the management of multiple cloud technologies over a distributed environment.

C2. Edge Service Orchestration

Edge service orchestration determines the optimal environment for running an edge application software to serve an end-user at a certain location and time. The orchestration may involve selecting the site (or sites) to deploy the application, setting up the optimum cloud environment in the site, or configuring the existing environment with specific characteristics.

The fundamental challenge with edge service orchestration refers to the optimal allocation of resources in real-time to deliver a seamless and efficient user experience, while at the same time maintaining the architectural flexibility of edge computing environments. The respective decision-making process is intricate, involving considerations of latency, available resources, geographical location, connectivity and more.

Edge service orchestration, as described by the GSMA Operator Platform Group [33] within the concept of Operator Platform, is paramount to provide the standardisation of a common federation interface that expands the operator's footprint. It allows edge providers to offer edge services to their customers directly or in collaboration with other providers when edge services need to be delivered beyond their footprint or their capabilities.

Existing solutions are either too centralised, lacking robustness and scalability, or too generic, failing to account for the unique requirements of telecommunication services. The orchestration capability must be capable of multi-dimensional decision-making that includes metrics related to hardware capabilities, network latency, and Service Level Agreements (SLAs).

Most of the solutions do not address the requirements related to site selection and traffic steering, as these are required to deliver and preserve Quality of Service (QoS) and performance, both at the computing and networking layers:

- Select the site(s) to deploy an edge application, considering the user(s) location, the application requirements, and the available computing and connectivity capabilities.
- Set up the optimum cloud environment and deploy the application instance in the selected site(s). The decision to deploy an application image on one or many sites may depend on the application design (e.g. redundancy to improve resiliency).
- Configure the user traffic to be optimally routed to the application in the selected site(s).
- In the case of mobile users, execute an optimal selection of the user plane attachment point (e.g. UPF/PGW configuration for data entry point).
- Request a specific QoS for the edge connection following a certain service level required by the edge application.
- Configure the entry point to the edge application in the device client application (e.g. IP address of the edge application server and attachment to specific network slice to connect to that server).
- Start/stop the on-line billing process (e.g. OCS) based on the subscribed billing scheme for the specific applications.

All the aforementioned actions need not only to be performed but must also be performed in a zero-touch way through an intent-based interface. Multi-tenancy should also be facilitated, i.e. preserving the integrity of each tenant's environment in the selected site(s). Furthermore, application-aware resource allocation needs to be ensured, with site selection based on the functional and technical requirements of each application (e.g. expected latency, compute resources demanded) and different criteria (e.g. closest site to reduce latency, site where more compute resources are available for load balancing, best performing site, or the cheapest for the required processing).

C3. Mobility Management

Ensuring edge service continuity and maintaining the quality of service are essential aspects, and become especially complex in edge applications for mobile devices, as computing and connectivity conditions may change as the user moves. The support for mobility across a multi-provider, multi-market distributed computing environment is partially addressed in the previous initiatives. It will require a smarter integration with the network and solutions to run the aforementioned edge service orchestration by dynamically reselecting, if needed, the edge computing node serving a

certain customer while moving. This also applies to roaming scenarios in which the end user moves to a visited[9] network.

To manage user mobility and adapt edge service delivery accordingly to preserve quality of service, the solution needs to:

- Periodically monitor the user location (and user service experience) to check if an edge node reselection is needed and, as a result, an edge application instance migration.
- Configure the user traffic to be routed to the application instance in the new serving edge node.
- Execute data plane attachment point (re)selection for service continuity to break out closer to the new edge node (e.g. UPF/PGW configuration for data entry point).
- If needed, interact again with network to request specific QoS in order to keep the service level agreements defined for an edge application in the connection with the new edge node.

Even if efforts are made in 3GPP and ETSI with architectural and testing work (e.g. ETSI MEC4AUTO group in collaboration with 5GAA), and APIs like "edge discovery" or "traffic influence" are being defined in the Edge Cloud WG at CAMARA project, the mobility scenarios for edge applications are not completely defined, addressed, and solved in working solutions.

C4. Federation at Different Levels

Despite existing initiatives, insufficient solutions are available for the federation across multiple dimensions: network, cloud, data, and services. Such federation is crucial for ensuring seamless interoperability and scaling:

- At network level, a proper and stable interconnection between networks is required to provide the right levels of bandwidth, latency, and jitter for edge applications that connect users subscribed to different network operators. A representative example is the case of a V2X collision warning system alerting vehicles connected to different networks, or gaming sessions between users in different networks. A proper interconnection between edge providers (network operators) is equally important as a close and performing computing.
- At resource (cloud) level, an efficient connection of resource catalogues is fundamental for a holistic view of the available computing resources in the covered markets and to dynamically request the allocation or release of resources and the management of the edge applications' lifecycle.
- At service level, the necessary business relationships (service agreements, SLAs, billing and charging schemes) need to be set up between the federated edge providers so that each of them can act of behalf of the others to offer computing capacity over the combined footprint.

[9] A network is said to be "visited" when an end user connects to it and that end user is not subscribed to the operator that owns and operates it. In contrast, "home network" is the one the end user is subscribed to.

- In upper layers (like data and AI), additional federation mechanisms may be required but this is out of the scope of this paper.

Multi-cloud networking refers to the practice of interconnecting multiple cloud computing environments or platforms, such as public cloud providers (e.g. AWS, Azure) and private clouds, to create a unified and interconnected computing infrastructure. Various network connectivity options are utilised to implement multi-cloud networking, for instance virtual private networks (VPNs), direct private peering, software-defined wide area networks (SD-WAN), or public cloud interconnectivity solutions like AWS Direct Connect or Azure ExpressRoute. Nevertheless, none of them provides a solution to interconnect different edge service providers while preserving the required latency and bandwidth.

Early implementations of edge cloud are being deployed in a highly varied manner across different operators. At present, there is no common blueprint outlining how network and service components should be integrated to deliver edge services. The absence of standardised blueprints necessitates additional network testing and integration, leading to increased costs for overall solutions and an extended Time To Market (TTM).

There is an insufficient integration of network enablers with cloud services, which are needed to provide value-added services such as on-demand dedicated quality. Each telecommunications operator implements its own model of network integration, which has limited reusability in other networks. CAMARA has taken a step forward with the development of the "edge discovery" and "traffic influence" APIs [34] that allow the selection of the optimal edge node to serve a certain customer in a certain location and re-route the traffic optimally to that node, but additional control on the connectivity and QoS is required.

Although there have been several trials regarding resource and service federation [35], [36] and work is ongoing at GSMA [37], there is in general no solution being developed or designed that covers all the different federation levels. This is fundamental to deliver a universal service across different edge operators.

C5. Capability Exposure Functions

Advancements have been made in exposing network functionalities via APIs, but complete and standardised exposure functions for the lifecycle management of edge computing resources and edge applications are lacking. This gap hampers the development of new services and applications that can fully leverage the Telco's capabilities.

The CAMARA project has created an edge cloud repository for edge-related APIs [38], but most of the work so far focuses on edge connectivity (edge discovery and traffic influence). The work on edge resource and application lifecycle management APIs is ongoing.

C6. Security and Compliance

Attention to digital sovereignty, systematic risk, economic risk, related technical, organisational and operational security measures, dynamic assurance, and continuous monitoring is further increasing. However, the current frameworks may not sufficiently meet all the complex security and regulatory compliance requirements specific to edge cloud continuum operations in multi-tenant and multi-actor infrastructures and related environments. In this context, the Digital Decade 2023 policy programme focused on delivering the risk-based and even "all-hazard approach" strategies such as the Cyber Security Strategy [39]. In addition, new or upcoming European regulations that are or can be within the scope of these edge cloud ecosystems should be considered, including already applicable regulations. These regulations refer to GDPR (EU 2016/679) and ePrivacy Directive (2022/58 EC), the latter of which is still under revision to become the ePrivacy Regulation [40], as well as to others such as the NIS2 Directive (EU 2022/2555) (NIS2) [41], the Resilience of Critical Entities Directive (EU 2022/2557) (CER) [42], the Cyber Resilience Act (CRA) [43], the Cybersecurity Act (EU 2019/881) (CSA) [44], the Cyber Solidarity Act [45] (which is currently a proposal), and the Artificial Intelligence Act (AI Act) [46].

One dimension of the multi-dimensional challenges is the supply chain or in general in stakeholder ecosystems, the (co-) accountability, governance and management, including maintaining a secure and compliant multi-tenant, multi-actor environment, especially when the underlying infrastructure is shared among different organisations and influenced by various and numerous stakeholders with possibly different values and interests.

C7. Disaggregated RAN hardware acceleration

Open RAN has specified a functional architecture where the Radio Access Network (RAN) layers and functions are split across three major components: the Radio Unit (RU), the Distributed Unit (DU), and the Centralised Unit (CU) [47]. CU deals with higher layers (L3) of the 5G RAN protocol stack and can be implemented on server platforms without relying on specialised hardware, hosted typically at the Near Edge or the Cloud. On the other hand, Layer 1, the physical layer signal processing functions, executes at the RU and the DU, typically at the Far Edge.

5G RAN needs to meet requirements in terms of latency, capacity, and energy efficiency, with the highly compute-intensive L1 functions having the tightest time budget. Specialised hardware is deemed mandatory to meet the RAN's performance, power, and cost requirements. In addition to the RAN itself, low latency AI-native applications are expected to be deployed at the Edge to address end-user AI-based services.

The term *hardware accelerator* refers to specialised hardware on which the most compute-intensive signal processing can be offloaded. Examples of HAs include ASIC, FPGA, DSPs, and GPUs. While Field Programmable Gate Arrays (FPGA) was a leading candidate until recently, mainly for its re-

programmability, currently vendors are proposing two types of architectures based on Application-Specific Integrated Circuit (ASICs) and System-on-Chips. The first is the so called "Look-aside". L1 functionality is split between server CPUs and HAs, where the HA may be housed on a separate PCIe card or integrated on-die with the CPU's cores. Chip manufacturers such as Intel promote this latter approach to remove the need for extra PCIe cards (to reduce power consumption and supplementary data movements). The second, referred to as "inline", shifts L1 processing from CPU to specialised System-on-chips, connected to servers through PCIe. The downside of inline acceleration with customised silicon is that it is programmed with proprietary tools (or is only partially programmable).

Representative challenges include:

- Avoiding vendor lock-in through adherence to O-RAN Acceleration Abstraction Layer that specifies a common and consistent set of interfaces to different hardware components.
- Disaggregation comes with complexity, with increasing interoperability issues, thus requiring intensive testing and certification.
- Anticipation of scalability requirements at the far edge in view of centralised DUs aggregating multiple radio sites.
- Provision of sufficient programmability, to be future-proof regarding evolving standards and new algorithms, and to allow resources to be reused by other applications when radio traffic load is significantly lower than capacity. Fixed-function ASICs, whether embedded or on separate boards, cannot be re-purposed.
- Enabling flexible and dynamic selection of compute devices based on current or predicted traffic and workload.
- Economies of scale can be achieved if pooling gains can be leveraged. Not only must the physical infrastructure be "poolable" but intelligent scheduling is required to use resources optimally between L1 processing and AI/ML algorithms that are expected to run at edge sites.
- Analyse trade-offs between "one size fits all" versus specific hardware infrastructure based on deployment needs.

SECTION 4: PROPOSED ACTIONS AND RECOMMENDATIONS

Following the challenges identified in the previous section, the current section provides a set of actions and recommendations.

S1. Multi-provider Container Cluster Manager

The definition of Container Cluster Manager (CCM) in the last version of ETSI NFV [48] refers to a component that can manage the setup, configuration, and monitoring of multiple workload clusters, which is required in both distributed environments and those related to the deployment of edge applications, but this concept should be extended to become multi-cloud. This evolution to multi-cloud cluster management (MCM) implies managing multiple clusters that may use different container cluster technology (i.e. different k8s distributions) because it is unlikely that the underlying infrastructure for a distributed and federated edge architecture will be based on a single container cluster technology. Most probably, multiple providers will coexist and, to simplify operation, all of them must be managed from one single cluster manager, facilitating the edge service orchestration (*Challenge C1*).

S2. Open source Edge Service Orchestration

Given that there is no active initiative to build a reference implementation of the multi-cloud and edge service orchestrations (*Challenges C1 and C2*), an open source community should be promoted to work collaboratively on a first complete open source implementation that can used as a reference for commercial edge orchestration solutions. This community must identify which existing assets (products, open source code, etc.) can be leveraged and which code needs to be developed, to allow a multi-cloud and vendor-agnostic capability and define the interfaces required for communication based on the work done in other existing initiatives such as Sylva, CAMARA, or NEPHIO.

S3. Intelligent Edge Resource Matching Algorithm

To tackle the challenge of multi-tenancy and application-aware dynamic resource allocation in the edge site selection process (*Challenges C2 and C3*), an intelligent edge resource matching algorithm can be researched, developed, and implemented. This algorithm would consider several parameters such as expected latency, available compute resources, and cost factors, and would prioritise them based on the application's requirements. Additionally, ML/AI models could be

utilised to predict future resource requirements based on historical data, aiding in more intelligent and effective decision-making in terms of edge resource planning or predictive maintenance.

S4. Policy-Driven Orchestration

To ensure service continuity, especially in edge applications for mobile devices (*Challenge C3*), a policy-driven orchestration approach can be taken. Policies for each type of service could be defined and embedded into the orchestration process. These policies would define the rules for maintaining service quality, setting up failover procedures, and re-routing traffic as needed.

The policies should consider the fulfilment of application, functional and technical requirements (e.g. latency, compute requirements) and other criteria (e.g. load balancing, best performance, or cost). AI may be used to dynamically change service configuration and resource allocation to ensure the edge behaves according to the policies.

Policies are also relevant at resource level in the edge federation interfaces (*Challenge C4*).

S5. Open source Federation Manager

Operators should complete the standardisation work started in GSMA and ETSI and develop an open reference implementation of the federation management modules to facilitate interoperability between different edge compute platforms. Launching an open source community could help in accelerating the availability of these open federation mechanisms in existing and new edge compute platforms by combining efforts from operators and software companies in this endeavour (*Challenge C4*).

S6. Federated Marketplace for Cloud-Edge APIs

A federated marketplace for cloud and edge services would foster interoperability and portability and act as a pivotal facilitator, expediting the adoption of cloud and edge service offerings across Europe (*Challenge C4*). These offerings would adhere to EU regulations and meet European standards and user expectations in terms of reliability, security, portability, and energy efficiency. This marketplace is essentially a catalogue of APIs, encompassing both cloud and network Service APIs that augment edge cloud services. The marketplace will feature APIs from existing initiatives like GSMA and CAMARA, as well as those developed during the RDI phase of various projects within the IPCEI-CIS.

The marketplace may be built upon concepts like the Telco Finder defined by GSMA at the Open Gateway initiative.

S7. Network Integration Blueprint

Operators should collaborate in the design, integration, and testing of open reference implementations (blueprints) for the integration of network and service (computing) components at the edge nodes, and for the adaptation of transport and core networks to facilitate the delivery of edge application requirements such as latency, jitter, or bandwidth. The aforementioned integration and adaptation would facilitate a uniform customer experience across a multi-provider federated footprint (*Challenge C4*) and enable economies of scale in the solution.

Initiatives like IPCEI-CIS can host this kind of collaboration.

S8. Standard Edge Computing APIs

Operators should harmonise the way to access edge and cloud resources and services from multiple technologies and providers, to facilitate aspects like edge service orchestration (*Challenge C2*) and enable a uniform experience for edge API consumers regardless of the market where they run the applications or the operator that serves the APIs.

Projects like CAMARA are hosting this kind of collaboration at its Edge Cloud WG and have started producing APIs related to edge discovery and edge connectivity control, but should address *Challenge C5* and increase efforts to provide the standard open APIs to consume edge and cloud computing as a service. This must be done in close collaboration with the open source federation manager (as described in *S5* above), as federation (provider-facing) APIs and consumer-facing service APIs need to be coordinated.

S9. Attribute-Based Access Control & Compliance Monitoring

Challenge C6, maintaining a secure and compliant multi-tenant environment, is especially relevant when the underlying infrastructure is shared among different organisations.

Implementing Attribute-Based Access Control (ABAC) can provide fine-grained access control based on multiple attributes like role, location, and time. On the other hand, Compliance Monitoring solutions should continuously scan the environment against policy benchmarks, reporting and rectifying any deviations.

S10. Hardware Acceleration

One way to address the hardware acceleration challenges (*Challenge C7*) is to establish a liaison between the IPCEI on Microelectronics (MECT) and the IPCEI on Cloud (CIS) to align the requirements and facilitate cross-collaboration. The IPCEI MECT should consider, for instance, the

convenience of designing telco cloud-specific hardware accelerators that could also serve for other edge applications (AI/ML, video processing, etc.).

Additionally, the European Commission should continue to support the EU-based hardware acceleration providers participating in the IPCEI-MECT and promote the emergence of new ones.

Finally, the Open Programmable Infrastructure (OPI) project hosted by the Linux Foundation can help to address the hardware acceleration challenges and would benefit from more traction from the telecom solution providers and support from the European Commission.

SECTION 5: FUTURE TOPICS

This thematic roadmap focuses on the most prominent aspects for the telco edge computing development. Additional aspects (in the scope of a long-term time scale) include the following:

- The edge may host data and data management and control mechanisms for the data layer, becoming, for instance, part of the mechanism for data sharing (data spaces). Initiatives like Gaia-X, DSBA, or DSSC are a reference in this topic and their work with edge computing (and storage) should be assessed.
- On the application layer, the compliance of edge applications, including network functions, with the cloud-native principles is an important topic to address in the migration of applications to the cloud-edge and in the development of new edge applications. Projects like Sylva have started to address the testing and certification on this issue, but extensive work is required to ensure a true migration to the cloud.
- AI at the edge for telecom networks. AI/ML will be fundamental towards efficient allocation, use, and management of edge compute and network resources, from planning to upgrades and operation. On the other hand, the edge may be a key location to run parts of the AI/ML logic (e.g. pre-training, fine-tuning, RAG, inferencing).
- While this document is focused on the telco cloud, there may be specific requirements for other types of edge and clouds, and for other applications like V2X, Railways, or CDNs that are not specifically addressed.
- The tighter integration of network and cloud that is currently being designed for the 6G architecture in bodies like 3GPP was not included in the scope of this paper, but the initiatives described in this paper may contribute to this direction.
- Additional network functions that might be exposed by telecom operators, beyond the edge connectivity that was addressed in this paper, have not been discussed. These include the Network Exposure Function (NEF) that allows the configuration of the quality of service for a data flow, the Network Data Analytics Function (NWDAF) that can provide data about network status and performance, the PCF (Policy Control Function) that allows the setting of policies to control network behaviour, or the Network Slice Orchestrator that provides functionality to manage the lifecycle of network slices.

SECTION 6: CONCLUSIONS

Addressing the telco use case needs described in Section 2.1 is crucial to position Europe in a leading role in the global telecommunications landscape. By tackling challenges tied to complex and demanding use-cases, for example that of telecommunications networks, solutions that benefit a wide range of industrial applications can be developed. Furthermore, the scaling of cloud and edge infrastructures directly correlates with the European Commission's objectives for digital growth and sustainability.

Opening up telecommunication infrastructures for customer utilisation aligns closely with the European Commission's objectives for a Digital Single Market but presents notable challenges – as described in Section 3. This move democratises access to high-performance computing resources, fosters innovation, and catalyses economic growth. Moreover, it positions the European Union as a hub for technological advancement and sustainable digital ecosystems.

The realisation of the telco cloud in Europe should be based on the European principles of cooperation, coordination, and standardisation, which are the basis for both the construction and opening of infrastructure. The present document contributes to this.

The European Commission is strongly encouraged to continue and amplify its public policy and financial support to the already-existing initiatives listed in Section 4, and support the implementation of the recommendations formulated in Section 5 in a rapid and focused manner. .

REFERENCES

[1] Cloud Alliance, https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance

[2] European Industrial Technology Roadmap for the Next-generation Cloud-Edge, https://ec.europa.eu/newsroom/dae/redirection/document/102590

[3] White Paper - How to master Europe's digital infrastructure needs, https://digital-strategy.ec.europa.eu/en/library/white-paper-how-master-europes-digital-infrastructure-needs

[4] The Path to the Digital Decade, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en#the-path-to-the-digital-decade

[5] OpenNebula code repository, https://github.com/OpenNebula/one/

[6] OpenNebula Community Forum, https://forum.opennebula.io

[7] OpenNebula Open Cloud Reference Architecture, https://docs.opennebula.io/6.8/overview/cloud_architecture_and_design/open_cloud_reference_architecture.html

[8] OpenNebula Edge Cloud Architecture, https://docs.opennebula.io/6.8/overview/cloud_architecture_and_design/edge_cloud_reference_architecture.html

[9] ONEedge5G project, https://opennebula.io/innovation/oneedge5g/

[10] Anuket, https://anuket.io/

[11] Sylva, https://sylvaproject.org/

[12] Sylva technical principles architecture, https://sylvaproject.org/#architecture

[13] Nephio, https://wiki.nephio.org/display/HOME/Overview+of+Nephio

[14] Camara, https://camaraproject.org/

[15] OPI, https://opiproject.org/

[16] StarlingX, https://www.starlingx.io/

[17] ETSI NFV, https://www.etsi.org/technologies/nfv

[18] ETSI NFV Release 4, https://www.etsi.org/deliver/etsi_gs/NFV/001_099/006/04.04.01_60/gs_NFV006v040401p.pdf

[19] ETSI NFV Release 5, https://www.etsi.org/deliver/etsi_gr/NFV-IFA/001_099/046/05.01.01_60/gr_nfv-ifa046v050101p.pdf

[20] ETSI White Paper, "Evolving NFV towards the next decade", https://www.etsi.org/images/files/ETSIWhitePapers/ETSI-WP-54-Evolving_NFV_towards_the_next_decade.pdf

[21] OSM MANO, https://osm.etsi.org/wikipub/index.php/Release_notes_and_whitepapers

[22] Open Source MANO Working Procedures, https://portal.etsi.org/Portals/0/TBpages/OSM/Docs/OSM%20WP%20v1.0.pdf

[23] IPCEI-CIS announcement, https://www.bmwk.de/Redaktion/EN/Artikel/Industry/ipcei-cis.html

[24] EU Data Strategy 2020, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

[25] Manifesto for the development of the next generation cloud infrastructure and service capabilities in 2022, https://www.bmwk.de/Redaktion/EN/Downloads/M-O/manifesto-for-the-development-of-the-next-generation-cloud-infrastructure-services-capabilities-in-2022.pdf

[26] EU Green Deal, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en
[27] EU Industrial Strategy, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-industrial-strategy_en
[28] Europe's Digital Decade: digital targets for 2030, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en
[29] XCP-ng, https://www.xcp-ng.org
[30] XEN Orchestra, https://www.xen-orchestra.com
[31] ReNESENS project completed the porting of XCP-ng on the Kalray DPU architecture, www.kalrayinc.com
[32] CNCF Whitepaper, "Accelerating Cloud Native in Telco", https://www.cncf.io/blog/2023/12/13/announcing-the-accelerating-cloud-native-in-telco-whitepaper-v1/
[33] GSMA Operator Platform Group, https://www.gsma.com/futurenetworks/operator-platform-hp/
[34] Camara – EdgeCloud GitHub, https://github.com/camaraproject/EdgeCloud
[35] GSMA, TEC Pre-Commercial Trial – Edge Compute Service, https://www.gsma.com/get-involved/gsma-foundry/tec-pre-commercial-trial-edge-compute-service/
[36] Telefónica signs an agreement with KT Corp., China Unicom and Telstra to collaborate on the multi-operator Edge Computing experience, https://www.telefonica.com/en/communication-room/press-room/telefonica-signs-an-agreement-with-kt-corp-china-unicom-and-telstra-to-collaborate-on-the-multi-operator-edge-computing-experience/
[37] GSMA Operator Platform Group, Edge Federation APIs, https://www.gsma.com/futurenetworks/resources/gsma-operator-platform-group-east-westbound-interface-apis-version-3-0/
[38] CAMARA Edge Cloud API Repository, https://github.com/camaraproject/EdgeCloud
[39] Cybersecurity Strategy for the Digital Decade, https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0
[40] ePrivacy Regulation Proposal, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017PC0010
[41] NIS2, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2555&qid=1675242238329
[42] CER, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2557&qid=1673965247308
[43] CRA, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0454
[44] CSA, https://eur-lex.europa.eu/eli/reg/2019/881/oj
[45] Cyber Solidarity, https://digital-strategy.ec.europa.eu/en/library/proposed-regulation-cyber-solidarity-act
[46] AI Act, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
[47] O-RAN Reference Architecture, https://docs.o-ran-sc.org/en/latest/architecture/architecture.html
[48] ETSI GS NFV-IFA 036, https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/036/04.03.01_60/gs_NFV-IFA036v040301p.pdf