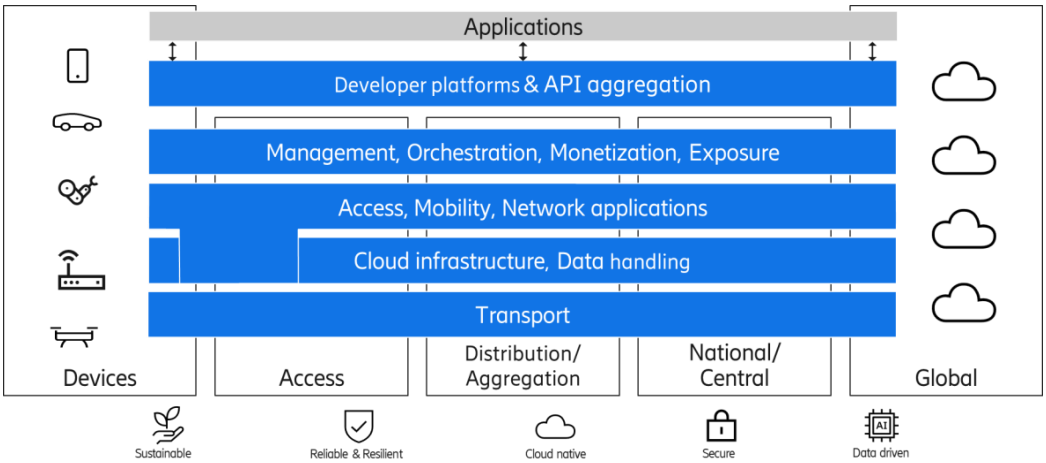
ERICSSON

Future Network Architecture 2026

Dependence on mobile networks is constantly increasing as their versatility enables support for society-, mission-, and business-critical applications. The network of the future will progress from current 5G through to 6G and maintain its position as the most expansive innovation platform observed. It will adopt a horizontal architectural paradigm and be sufficiently flexible to accommodate a greater variety of use cases than ever before.



150 years

Table of Contents

1 Executive summary

Mobile networks are critical for society and business moving from only MBB-connectivity to supporting efficient service introduction and efficient networks via programmable platforms able to expose capabilities via standardized APIs.

Future networks must meet increasing demands for coverage, efficiency, resilience, security, and advanced services in a cyber-physical continuum linking digital and physical systems.

The Ericsson Global Architecture (TGA) describes the view of the separation of the network into horizontal layers and vertical deployment locations. This modular, cloud-native design is intended to support openness, rapid service introduction, best performing network including security, sustainability, and heavy use of data and AI.

6G should be an evolution of 5G. The core network will be extended to support 6G radio, RAN will support a new 6G RAT as well as spectrum sharing(MRSS).

Several cross-cutting capabilities are highlighted:

- Autonomous networks: progressing toward "zero-touch" operations and rapid service and product introduction using intent-based management, AI/ML, and agentic AI.
- AI in the architecture: AI will be embedded throughout the network and in AI agents, coordinated and managed through emerging AgentOps practices.
- Exposure and APIs: a layered exposure model allows CSPs to offer network capabilities via standardized APIs with strong consent and data-privacy controls.
- Data handling: a federated data mesh approach (including Ericsson's Federated Data Lake) to collect once and reuse data across analytics, AI, automation, and exposure, while respecting governance and quality.

Mission-critical segments require very high reliability, availability, and resilience, as well as strong security and sometimes operation under degraded conditions. 5G/6G, together with features like non-terrestrial networks, integrated sensing and communication (ISAC), precise positioning, etc. will support these sectors and their evolving deployment models.

Enterprises are supported by multiple private and hybrid deployment options and stresses the need for a unified application experience across public and private domains.

Ericsson proposes a gradual, standards-aligned evolution from 5G to 6G, centered on cloud-native, modular architecture; pervasive AI and automation also meeting the need for rapid service introduction ; open and programmable interfaces; differentiated and dependable services; and robust security and resilience to support the digitalization of society and industry by 2030.

2 Introduction

The versatility and evolution of the mobile networks to develop new services and to support new industries is a major opportunity for communication service providers (CSPs) to leverage and satisfy different industries different digitalization needs.

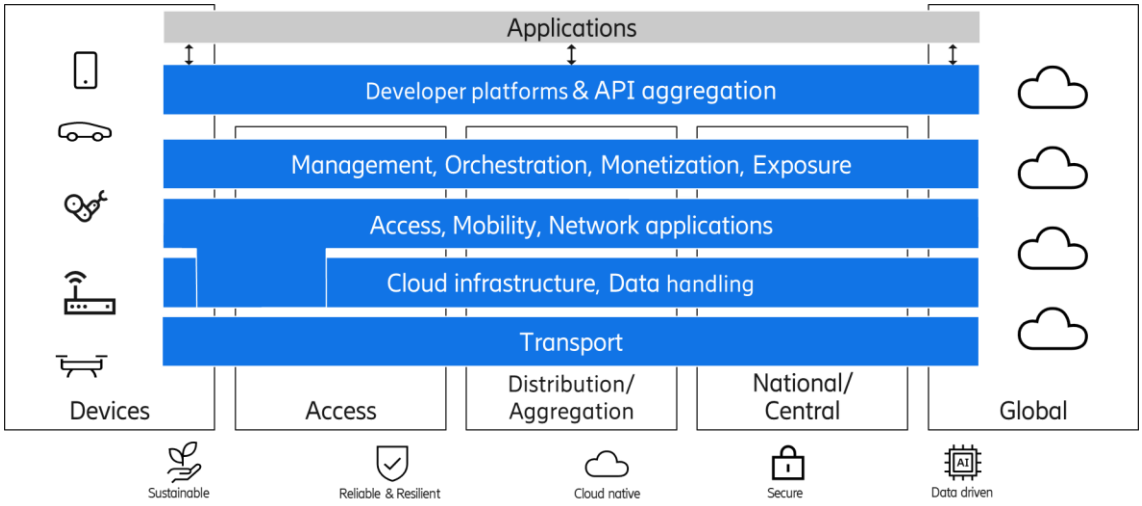


Figure 1 The Global Architecture

Today's mobile networks are undergoing a transformational phase, evolving into a major innovation platform that supports autonomy, programmability, and exposure of capabilities to a wide range of users.

A robust, well-considered network architecture is essential to address these diverse and expanding use cases. The horizontal structure of The Ericsson Global Architecture (TGA), Figure 1, is designed for fast introduction of new services and products meeting requirements for openness, programmability, exposure, and autonomy, and provides a unified conceptual framework applicable to Communication Service Providers, enterprises, and mission-critical networks.

Serving as a high-level view for architectural discussions, TGA offers a generic illustration that leads into more detailed analysis. Effective network architecture relies on narrow, stable components that break down complex problems, enabling orderly integration of new functions with minimal impact on existing components and layers.

The TGA consists of 6 horizontal domains and an Application domain external to the network:

- **Transport**
  Contains functionality for transmission and transport primarily between sites but also within some sites

- **Cloud infrastructure, Data handling**
  Contains functionality for secure processing and storage of both network functionality as well as application functionality. Data handling supports all network domains in collecting, storing, distributing and processing of data.

- **Access - Mobility - Network applications**
  Contains functionality securing all types of access as well as network integrated applications. Examples are CORE and RAN.

- **Management, Orchestration, Monetization, Exposure**
  Manages and controls the network e2e (RAN, Core, Transport & Infra domains), handles subscriber business management, exposes network functions to external applications, and will increasingly leverage AI-driven automation.

- **Developer Platforms and API aggregation**
  Developer platforms expose services and products, combining aggregated network services for developers and enterprises. API aggregation aggregates CSP APIs at scale and globally, exposing them for wholesale to external customers, typically developer platforms.

- **Applications**
  Contains network external applications interacting with the network and utilizing its exposed capabilities for communication, execution and storage.

Additionally the TGA has 5 vertical parts. **Devices** are end-user equipment or network setups established by users or enterprises outside the control of service providers; a device may also have an attached network and act as a router. **Access sites** are locations in wide-area or local dedicated networks that sit close to users. **Distribution or aggregation sites** are deployed to improve execution or transport efficiency, to enable local breakout, or to aggregate functionality for other reasons. **National or central sites** are typically centralized within a service provider's network or within a local dedicated (usually private) network. **Global sites** are centralized, publicly accessible locations—typically large data centers—that can be reached from anywhere.

Several important attributes are attached to the architecture such as Sustainability, Reliability & Resilience, Cloud Native principles, Security and being Data driven leveraging AI across the network.

Note that AI, though not explicit shown in the TGA, will be applied in all places of the network both to support intent driven automation and operations as well as increasing e.g. performance and adaptability in the networks to create new business opportunities.

3 Future Network Outlook – Trends and Drivers

3.1 High-level needs

From a future point of view, networks must be designed to include requirements for society- , mission-, and business critical networks as these increasingly rely on a digital infrastructure to be able to address the needs of the 2030 society through network capabilities, which in turn are delivered by technical solutions.

Four areas of drivers have been identified, addressing above needs and thus indicating the direction which future 6G networks should take.

The main driver for networks is to improve in coverage, efficiency, and trustworthiness aspects, as well as addressing new services and product introduction as well as application demands of novel advanced services; Inclusion & Integrity, Sustainability & Resilience, Security & Trust and Digitalization in order to meet application demands.

We expect mobile networks to be gradually upgraded to support 6G technology. Similarly to 5G which was both an improvement and expansion of 4G, it's likely that 6G will be a continued improvement of the "5G triangle" of eMBB, URLLC, mMTC services, while also adding new capabilities for delivering networking in the cyber-physical world , as shown in **Error! Reference source not found.**.

Integration of new capabilities expanding the network's scope includes; Compute, Artificial Intelligence (AI) and Sensing as examples.

This expanded framework, endorsed by ITU-R [1], represents a significant shift in network functionality. It moves beyond traditional communication services, positioning networks as multifaceted platforms capable of supporting a wide array of advanced applications.

The introduction of these new dimensions requires robust business and efficient mechanisms for market exposure thus ensuring that the technological advancements align with practical, market-driven needs.

The challenge lies in seamlessly integrating these new capabilities while maintaining and improving existing services. Such evolution aims to create a more versatile and powerful network infrastructure, capable of supporting the emerging technologies and use cases.

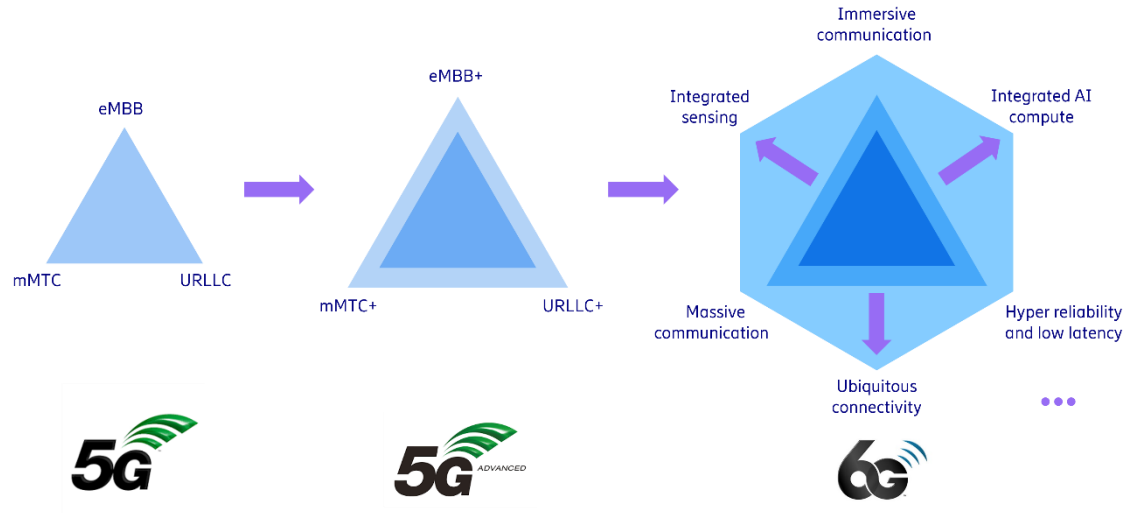


Figure 2 Evolving 5G and the journey to 6G: enhancing and expanding into new cyber-physical services

3.2 Major capabilities and use case trends

A number of use case examples have been identified for the 2030 networks stretching across Global Digitalization - Mixed reality - Autonomous Networks - Critical Services.

Increased requirements on the network capabilities can be identified, such as increased Up Link (UL) capacity, improved positioning accuracy and coverage, and improved support for service guarantees, etc.

Above use case examples in relation to the broader changes in the network are outlined below, referred to as networking and technology trends, needed to enable them. Existing 5G use cases, e.g. Fixed Wireless Access (FWA) will evolve into 6G.

3.2.1 Networking trends

1 - **Networks as platforms for connectivity and beyond**
Ericsson envisions networks evolving beyond connectivity into platforms that process data and offer advanced services—AI-driven sensing, localization, and reliable links—supporting applications for drones, digital twins, etc.

2 - **Performance differentiation**
While apps generally rely on best-effort mobile broadband, a growing demand can be seen for enhanced performance of future networks to provide dynamic APIs, robust SLAs, enabling mixed-reality integration.

3- **Data and AI everywhere**
AI, including AI Agents, will optimize networks and cut costs using e.g. on-network digital twins to enable predictive analytics or acting in intent-driven operations.

4 - **Diverse requirements, diverse deployments**
Future networks will be heterogeneous across technologies and deployments. Enterprises are expected to integrate private networks with public networks and public cloud to serve segments such as critical communications.

5 - **Resilient networks**
Network resilience and cybersecurity are critical as 6G must be able to handle increasingly challenging failure types to maintain seamless coverage and end-to-end service guarantees for critical services amid global challenges like climate change and geopolitical tensions.

6 - **Programmable Networks**
The transition to programmable networks [2], is driven by the demand for more versatile and customized products and services introduction. CSPs can leverage intelligent automation and real-time data for network optimization through Service Management and Orchestration Architecture. Intent-driven operations, powered by AI, will allow networks to autonomously execute desired outcomes, responding to dynamic consumer and enterprise demands.

3.2.2 Main Technology Trends

Above we listed network specific trends aimed towards 6G capabilities for use-cases in a cyber-physical world. In the following, we will also highlight several broader technological trends with implications on mobile networks and their architecture.

Horizontalization: Open interfaces now span multiple layers. Cloudification and multi-vendor vertical interfaces are essential as networks go cloud native introducing Kubernetes-based public and private infrastructure. Hybrid cloud alters security, delivery, and end-to-end NRAR(Network Reliability Availability and Resilience).

Softwarization: Telecom is software-first and programmable. Ericsson relies on extensive open source, making supply chain integrity and update practices vital. DevOps, CI/CD, telemetry and MLOps speed up time to market, though lifecycle management across diverse deployments remains challenging. Targeted observability and user-specific optimization matter. Specialized compute (GPUs, FPGAs, AI ASICs) pushes clouds toward real-time, deterministic behavior. AI and agentic LLMs will reshape API development and app creation.

End to end encryption: With 6G, most traffic and metadata will be encrypted, reducing network visibility. Apps will decide what data to share, requiring mutual value, especially for mixed reality and efficient eMBB which raises questions about user-plane security termination. Zero Trust will grow, with differing E2E expectations for enterprise control versus consumer privacy.

Resource efficiency: Sustainability is constraint and opportunity. Networks must support SDG (Sustainable Development Goals)-aligned use cases while minimizing spectrum, energy, materials and operations overhead. Deep sleep modes, legacy retirement, and early hardware/software choices are needed to meet performance with minimal power, including sensing and AI.

Digital twins: Digital Twins will anchor the cyber-physical world, mirroring networks at defined fidelity to predict impact, validate slices, improve energy efficiency, QoS, assurance, reliability and resilience. They reduce deployment risk by, e.g. offline verification. Device-level twins should be kept separate from network operations to avoid confusion.

Generative and agentic AI: GenAI moves into devices and apps, driving uplink need with local content creation which affects energy, data access and regulation. Agentic AI, agents that perceive, reason and act, will accelerate Autonomous Networks (intent-based control, predictive cross-domain optimization, etc.). Early value will be in customer care and operations expanding into embedded intelligence and AIaaS, lowering barriers for Hyperscalers and AI-native entrants.

Enterprise convergence: OT, IT and telecom now form networks shaped by as-a-service models. Enterprises embed connectivity into processes and favor standardized IT/web practices. NaaS and CPaaS shift toward outcome-based recurring services.

Related articles/additional reading:
Ericsson CTO Technology trends 2025

4 Network Capabilities

4.1 Autonomous Networks

The telecommunications sector is progressing towards fully autonomous, self-managing networks, termed "Self-x", that operate with minimal or no human intervention across all CSP's processes, supporting strategic objectives such as Zero-Touch, Zero-Wait, and Zero-Trouble operations and rapid service and product introduction. TM Forum facilitates this transformation by offering standardized architecture, open APIs, and methodologies (Autonomous Networks Framework and Manifesto), backed by major Communication Service Providers (CSPs) and vendors, to address escalating network complexity, reduce operational expenses, and accelerate advanced service delivery.

Industry stakeholders are advancing towards higher levels of network autonomy, leveraging TM Forum's Autonomous Network Levels (ANL 0–5). Presently, the sector averages an ANL of 1.9, indicating partial automation featuring basic closed-loop processes, but with ambitions to attaining Level 4. (ANL4 refers to a highly autonomous network which intelligently makes decisions, managing networks in complex scenarios, using predictive analytics while autonomously optimizes network performance and customer experience, adapting to various domains without manual help. See ref [4] for details).

This constitutes a pivotal shift for CSPs, demanding substantial enhancements across technology infrastructure, organizational structure, corporate culture, and data and AI capabilities. The adoption strategy focuses on targeted deployment of ANL4 within areas demonstrating the greatest return on investment, continuously optimizing high-value processes.

As automation evolves, operational priorities transition from manual planning and troubleshooting to automated, intent-driven charging and operational workflows, initially implemented in OSS/BSS and progressively expanding into core and RAN network functions. Realization of autonomous networks depends on the integration of several foundational technologies:

- Intent-Based Management: Interfaces enabling operators to articulate business intents at a high level rather than issuing specific commands.
- Artificial Intelligence (AI) and Machine Learning (ML): Core analytical engines that process real-time telemetry, support learning and reasoning, and facilitate decision-making without explicit programming.
- Agentic AI: Collaborative autonomous software agents responsible for predictive optimization, fault mitigation, and comprehensive service lifecycle management.
- Closed-Loop Automation: A continuous observe–analyze–decide–act cycle transitioning the network from reactive to proactive operations.
- Knowledge Management: Systematic approaches to producing, sharing, utilizing, and governing organizational knowledge and data assets.
- Data Management & Observability: Robust, accurate telemetry serving as the authoritative data source for the closed-loop observation phase.

TM Forum has articulated a clear vision and roadmap for autonomous networks, providing guidance on Autonomous Domains (ADs) and protocols within the Open Digital

Architecture (ODA). The TM Forum view can be seen as a top-down view describing the overall transformation of the CSP operations towards autonomous networks. The bottom-up view must also be considered, including integrating with standards and blueprints from other telecommunication bodies more focused on the network functionality, such as 3GPP and the Open-RAN Alliance (ORAN). In this area network vendor innovation, both based on and outside of standardization in automating the network functionality and its handling is important.

Ericsson's management and orchestration model separates functional management (services and function oversight) from realization management (software and infrastructure administration), optimizing cross-network management for cloud-native environments. This approach encompasses RAN, transport, core, E2E services, and experience management, all underpinned by unified enablers for data, AI, agent-based systems, and digital twins as depicted below in Figure 3.

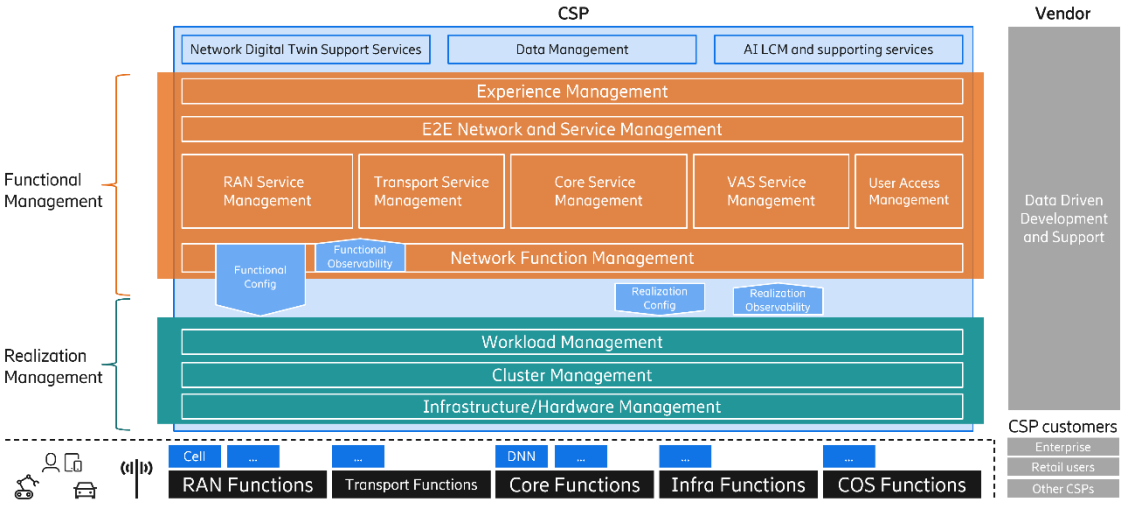


Figure 3 Ericsson target management and orchestration architecture

Autonomous Domains are fundamental constructs within autonomous networks, defined both operationally and architecturally/ functionally. Operational ADs encompass all lifecycle management tasks aligned to specific operational goals, whereas functional domains aggregate network elements and requisite LCM activities within their clearly demarcated boundaries. These domains often overlap, may be hierarchical, and can contain subsets of each other. Domain delineation is context-dependent; for example, enterprise deployments and CSPs may adopt distinct domain definitions according to their operational requirements.

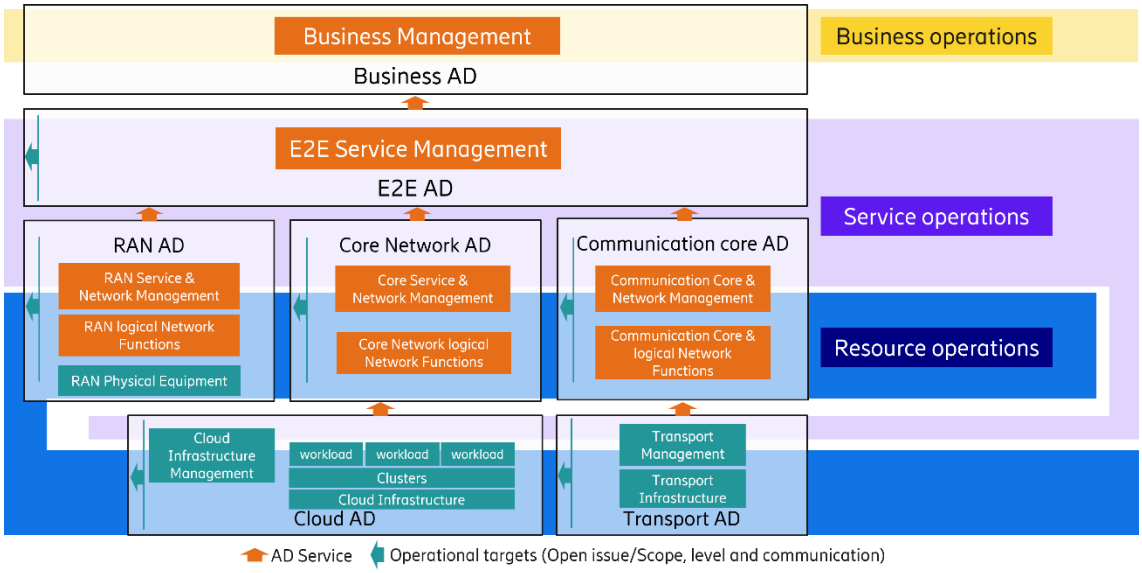


Figure 4 High level AN architecture blueprint for a CSP

Critical components of the high-level architectural blueprint include:

- Operational ADs deliver both resource-facing (RFS) and customer-facing (CFS) services, encompassing comprehensive resource and service operations.
- Full alignment with TM Forum architectures and guiding principles.
- Incorporation of best practices and influences from multiple industry standards bodies.

Related articles/additional reading:
Autonomy by Design
Intent-driven networks is a key step in the journey to autonomous networks
Intent-based networks in telecom operations - Ericsson

4.2 AI in the Network architecture

AI is not confined to a single mobile-generation nor to specific parts of the mobile network but is already present in 4G and 5G functions and widely seen as a key enabler for 6G, contributing to the architecture of future networks. Ericsson regards AI primarily as an implementation technology rather than a separate functional layer, advising caution against over-specification that may lock the industry into immature choices. The pragmatic recommendation is to reuse the 5G Core as a baseline for 6G and incrementally add agent capabilities and expose application-specific agent support in enablement and exposure layers rather than embedding it deep in Core or RAN functions. Standardizing essential open interfaces for multi-vendor interoperability is critical to avoid vendor lock-in while preserving room for continued innovation. Figure 5 illustrates how AI agents are added to a system.

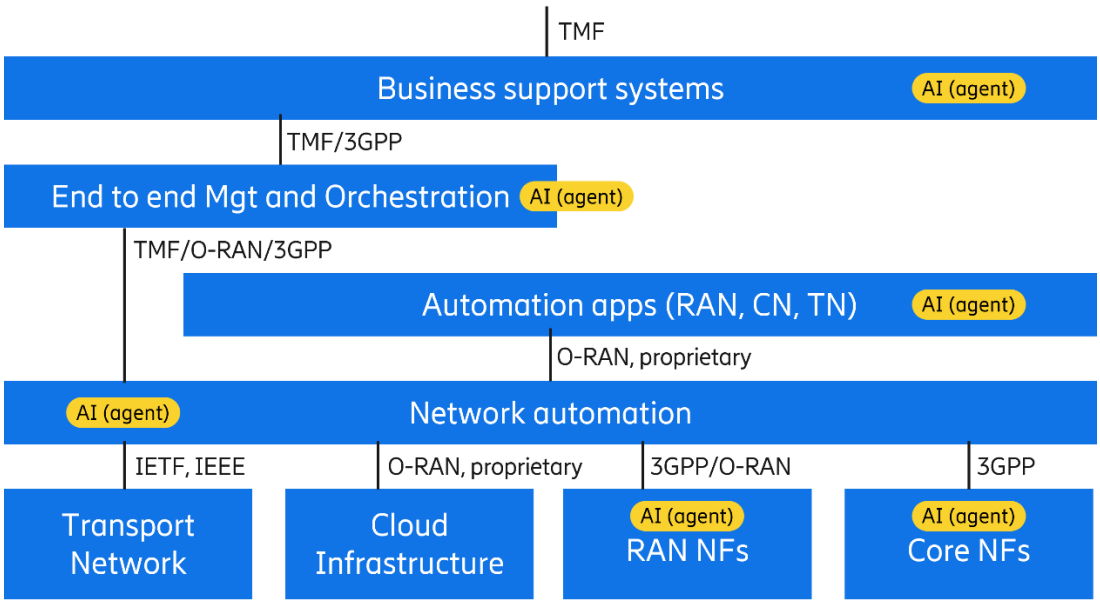


Figure 5 Agents as system add-ons

Standardization activity is already underway across multiple SDOs (Standard Development Organizations) and forums with discussions centered on AI/ML workflow handling, enabling/supporting functionality, APIs, and interoperability.

Agentic AI standardization has started but remains fragmented. TM Forum is leading with architectures and APIs for autonomous networks, including assets around Agentic AI protocols; 3GPP is exploring AI agents in its 6G studies; and O-RAN is preparing for agentic concepts within the SMO. Alignment, especially toward 3GPP, is seen as essential to achieve coherent cross-domain standards.

Agentic AI refers to autonomous systems authorized to act, decide, and self-initiate tasks on behalf of an entity. Agents perceive environments via sensors, protocols, and data streams, then apply rules, logic, or learned models to generate outputs, take actions, call tools, or execute code. They operate at runtime, maintain and update internal memory, and can collaborate via agent-to-agent communication. Behavior ranges from deterministic rule sets to Turing-complete reasoning, with machine learning used to adapt internal knowledge and improve decision-making over time.

The evolution from single-agent architectures to multi-agent systems (MAS) coordinated by an Agent Fabric is central to the telecom vision. An Agent Fabric provides discovery, communication, orchestration, and governance for many cooperating agents. Early, high-value use cases include intent management functions that close control loops autonomously and interact with other agents to fulfill intents, AI/ML-driven rApps in the SMO that automate complex RAN tasks, end-to-end service orchestration, troubleshooting and assurance workflows, and business operations. Integration with digital twins and predictive analytics can further reduce manual intervention and accelerate innovation towards Autonomous Networks, see also Autonomous Networks.

Operationally, AgentOps is emerging as the complement to MLOps/LLMOps. While MLOps focuses on model lifecycle like training, deployment, monitoring, and governance. AgentOps addresses runtime governance, agent deployment and scaling, tool usage,

chaining of agent behaviors, and multi-agent coordination. For telecoms, AgentOps will be critical to deploy agentic systems safely and at scale.

Agentic AI is poised to accelerate autonomy across network operations and service orchestration, reduce human intervention, and improve decision quality across planning, assurance, and optimization. Realizing this vision requires modular placement of agent capabilities, standard open interfaces for interoperability, cautious standardization to avoid premature constraints, and an agentic framework tuned to telco requirements so the ecosystem can grow in a safe, scalable way.

To support end-user AI Agent service, different example scenarios can be considered in mobile networks all requiring a balance between innovation and architectural feasibility:

- Maintaining the status quo with existing UE controls and APIs relying on use of Differentiated connectivity with enhancements such as enhanced uplink and definition of performance levels
- Adding Model Context Protocol (MCP) exposing CAMARA and OpenGW APIs for improved end-user AI Agent interaction
- Introducing new capacities or services such as network exposed information on ego-position, relevant sensing information or spatial data access
- Creating a general agent support framework for discovery and authentication for end-user AI Agents connecting to other AI Agents or to network resources such as exposed data
- Deep integration between End-user and Network-based AI agents.

The first option offers network evolution to handle AI-based traffic towards OTT termination, and the last may demand radical redesign of networks. The three scenarios adding MCP, new network capacities or even a general agent support framework could be the choice if CSPs are to take a stronger position in their ability to enable End-user agentic applications without disrupting 5GC network architecture.

Ericsson advocates pushing end-user AI agent support to the developer platform and exposure layers rather than embedding it deeply in the Core or RAN in order to minimize risk, accelerate market readiness, and support functionalities like QoS, collaborative services, and advanced analytics, ensuring networks remain adaptable to the growing demand for agent-driven applications.

Related articles/Additional reading:
AI agents in the telecommunication network architecture
Trustworthy AI - What it means for telecom
Defining AI native: A key enabler for advanced intelligent telecom networks

4.3 Dependable Networks

A dependable network provides connectivity services with binding conditions, for both parties (i.e., the service provider and the service user). These conditions may be defined using service-level objectives (SLOs), which may be used as the basis for a service-level agreement (SLA) between provider and user. An SLO may refer to QoS, availability (e.g.,

Communication Service Availability – CSA), or reliability (e.g., Communication Service Reliability – CSR) as examples.

The awareness of our dependency to networks is growing amongst enterprises, governments, and private citizens alike. The geopolitical tensions including national crises have created attention of the need of higher levels of NRAR. Today's networks are still susceptible to disturbance by e.g. natural disasters forcing CSPs to special measures of which many could be addressed by improvements to the network itself.

This presents an opportunity for CSPs to further monetize their networks (current and future) by differentiated connectivity, see chapters E2E Service Exposure and Differentiated , addressing a larger market of new customers with requirements beyond best effort connectivity.

5G standards already provide a good basis for realizing dependable networks, especially in the domain of time critical communications.

Another related area regards deployment practices where cost efficiency must be considered, e.g. what deployment is required to support a certain level of dependability; what level of dependability is possible with a given deployment, etc.

Related articles/Additional reading:
Dependable Networks – from best-effort to guaranteed performance

4.4 E2E Service Exposure

The services and capabilities of the mobile network platform is evolving to become a natural, digitally integrated asset used globally across industries and domains. This enables new digital business and value-creation opportunities, delivering benefits to society, and provide services to end users.

To enable this vision, the standardization must harmonize with exposure and business frameworks, policies, APIs and services, including global reach providing ease of consumption by developers in different roles, development stages and across several industry verticals, addressing both wide area and private networks, and devices, see Figure 6.

Network API Exposure has been available since 4G and is expected to become a major enabler in 5G as well as in 6G.

Ericsson (with other industry players) supports a new and more agile and market focused initiative for C/L1, driven in the context of GSMA (e.g. Open Gateway), Linux Foundation (Camara opensource project) and TMF all related to, but separated from 3GPP standardization.

The necessary revitalization of mobile connectivity includes differentiation, predictable performance and new business models to serve the "app economy" with new capabilities for the CSPs to be able to create new products with specific characteristics improving both existing applications and enabling new applications.

Figure 6 below is an overview of a layered exposure architecture, depicting an enterprise either consuming APIs exposed directly by a CSP or higher-level APIs exposed by a developer platform interacting with an API aggregator. Network capabilities can also interact via NNI exposed interfaces, e.g. for roaming scenarios. Interactions between networks can also be executed via federation mechanisms on the exposure layer.

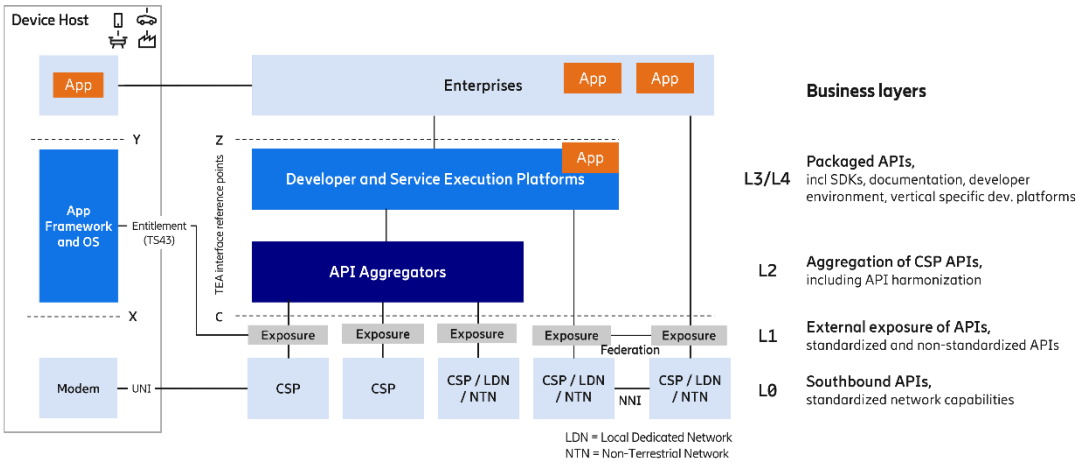


Figure 6 The exposure layered model

Each layer implements the services provided by the interface (and exposed by the APIs) and any other support functions, i.e. the layer includes the functional and non-functional SW to realize this layer:

Network technical capabilities APIs (L0)
Implements and exposes technical capabilities of the network, from the different domains: from 3GPP Core/RAN and OSS/BSS, via proprietary or TMF standardized APIs. L0 capabilities are not prepared to be consumed externally.

Network External Exposure (L1)
Implements and exposes network services from CSP Public Network or Local Dedicated Network to external customers, to either Application Service Providers, Enterprises, Developer Platforms or Aggregators and is responsible for data privacy management/consent management, API security and identity translation.

API aggregators (L2)
Implements functionality for API aggregation across a large number of CSPs to expose network API services in an aggregated form in a wholesale model to external customers, typically developer platforms. APIs exposed are Layer 1 standardized APIs (preferred).

Developer Platforms (L3/L4)
Implements functionality as part of the developer platform providers to expose services and products utilizing and combining, if needed, aggregated network services to its customers. Examples are combinations of Services and APIs streamlined for e.g. Fleet Management, Smart Industries, etc.

The Exposure platform, Figure 7, shows the network services exposed from a CSP Public Network or Local Dedicated Network to external customers. It provides the exposure platform for Network APIs and is responsible for the Network APIs composition based on API transformation (as referred by GSMA Open Gateway). It exposes them through standardized (preferred) interfaces, like CAMARA and TMF.

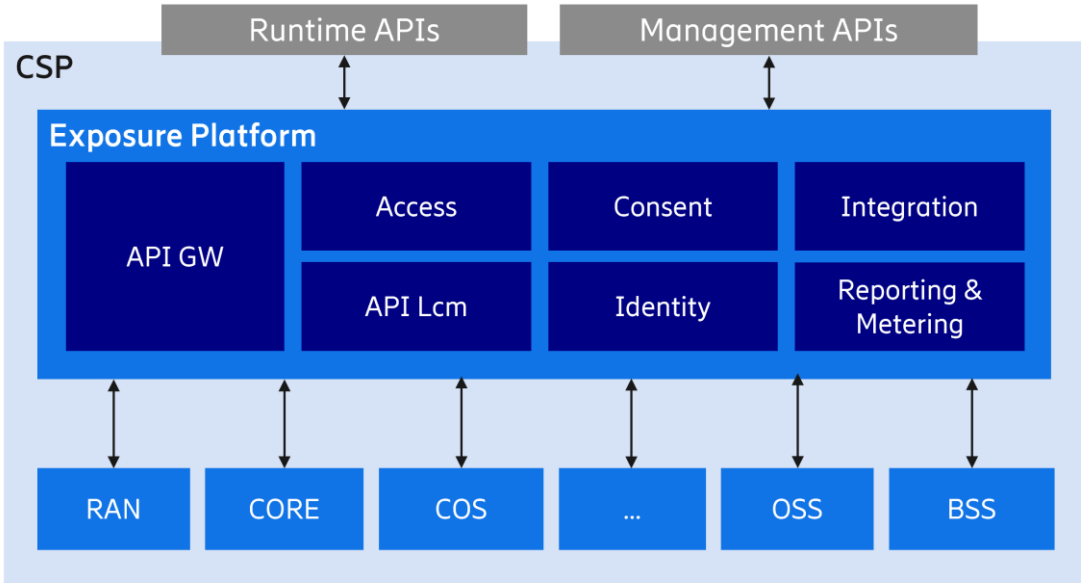


Figure 7 Exposure Architecture at CSP for Levels 0 and 1

One important area concerns data privacy management, a.k.a. Consent Management, which is a mandatory component of a service exposure platform at a CSP, allowing an operator to comply to legal requirements when processing and exposing privacy sensitive data of their subscribers respectively devices owned by their subscribers or other resources.

Of course Consent Management must be possible in a "roaming-capable API" scenario where enterprise devices (employee or IoT devices) roaming.

4.5 Differentiated Connectivity

The Ericsson vision for Differentiated Connectivity (DiffConn) aims to turn 5G SA networks into an innovation platform with new monetization models, and incentives for further network investment going beyond best-effort mobile broadband meeting needs of both new applications and new functionality in existing applications and enabling future use-cases. DiffConn is starting to happen in 5G SA, and it is important to secure smooth evolution towards 6G.

The goal is to create a healthier ecosystem where connectivity is monetized based on the value it delivers. Ericsson's vision is to create an open ecosystem where Application Service Providers (ASPs) can innovate on top of new network capabilities and CSPs monetize these capabilities.

Existing CSP networks need to evolve to support DiffConn with predictable, well-defined Performance Levels (PLs) which are defined using metrics like uplink/downlink bandwidth

and latency, coverage area, and availability, and can be part of SLAs with observability of SLA fulfillment.

The well-defined PLs enable CSPs to differentiate connectivity services and introduce new business models. DiffConn will work for all CSP customer types and even non-CSP deployments and support current business models: CSPs selling PLs directly (B2C, B2B) or via applications (B2B2X), while staying open to future models.

The e2e ecosystem for DiffConn includes multiple players that all need to contribute to and see the value of DiffConn. Examples of these are ASPs, CSPs, device/UE OS vendors, device/UE modem/chipset vendors, end users, enterprise IT administrators, aggregators, developer platforms, network vendors and regulators.

The DiffConn vision is realized by a combination of Subscribed QoS and Application-activated QoS activation models, creating an optimal balance between the two. In Subscribed QoS the DiffConn support is included e.g. in a monthly subscription between the subscriber and the CSP. Already today it is possible to leverage Subscribed QoS, thus preparing for the shift to Application-activated QoS, and ensure new slicing-based subscription models to support the long-term target.

In the Application-activated QoS model, DiffConn is activated from the application to the CSP via Network APIs. To really enable ASP innovation the ASPs must be involved in selecting and activating PLs, via network APIs for dynamic access, based on the needs of their applications. Applications also decide on mapping of their applications and application flows to the right PLs. Enabling exclusive usage of a PL for an application maximizes the potential value of the PLs for the ASPs and their users. This means that a PL is only used for a specific application and its application flows in a device in the CSP network.

Several granularity options will be available in the above activation models: whole device, application category, traffic category, application and application flow.

Figure 8 illustrates the high-level functional architecture of ecosystem entities involved in creating, setting up and managing differentiated connectivity service offerings. The main parts are the application client and server, the CSP's mobile network, and the aggregator / developer platform offering network APIs and the UE.

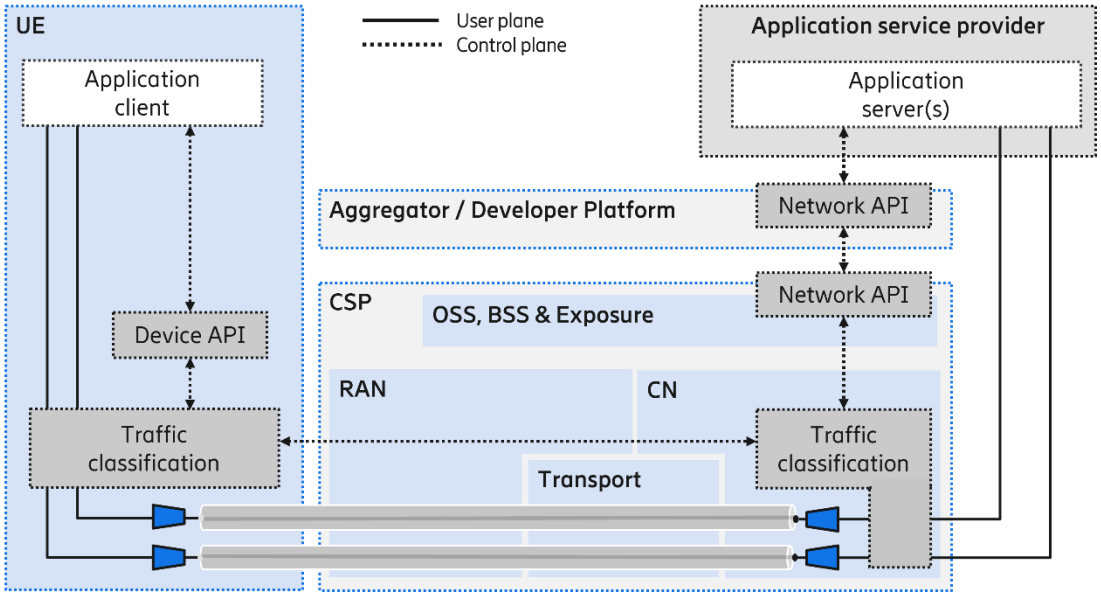


Figure 8 Functional architecture for Differentiated Connectivity

Related articles/additional reading:
[Differentiated connectivity: Unleashing the full potential of 5G]

4.6 Security

Security has historically implied that (enterprise) networks have been built to employ security control on its perimeters to prevent attacks. Today's networks should be regarded as untrusted and correct security measures must be deployed to protect the users (subjects) and resources that reside on these networks. In essence adopting Zero Trust Architecture (ZTA) principles.

The 6G security architecture will be an evolution of the 5G security architecture, with new security controls to fulfil the new 6G use cases. 5G has a solid security architecture where security in most parts is built-into the network but in some areas, they have been add-ons. It's assumed that the 6G architecture will use new protocols which should leverage the built-in security.

ZTA follows the principles, published by NIST in 2020 [3], which dictates that no network user, packet, interface, or device can be assumed to be trusted. Thus, implementation of ZTA affects all assets, including digital systems, people, and processes.

Even if 6G would be ZTA out of the box, perimeter protection like "gateways" between the domains will be needed to mitigate against security risks like lateral movement. ZTA is not replacing perimeter protection but is a complement to it.

The emerging area of exposure services has opened up new attack surfaces with increased security risks. Best practices exist today for protection of APIs and web applications, but the topic needs to be addressed further as well as in 6G.

Another topic of concern is Post Quantum Cryptography(PQC) whereby quantum computing is used to attempt to exploit vulnerabilities in the cryptography of Telecom

networks. With the standardization of the first quantum safe crypto algorithms already available from NIST, and work started in IETF to adopt these for coming IPSec and TLS versions these, older algorithms can soon be phased out. NIST and some regulators over the world are providing guidance when PQC crypto shall be introduced and non-PQC shall be phased out.

4.7 Positioning

5G integrates Location Services (LCS) for commercial and regulatory use leveraging how 3GPP defines nodes with various roles: Gateway Mobile Location Centre/Location Retrieval Function (GMLC/LRF), Unified Data Management (UDM) for subscription, Access and Mobility Management Function (AMF) for request processing, Location Management Function (LMF) for positioning calculations, Network Exposure Function/Common API Framework (NEF/CAPIF) for service exposure, and both User Equipment/Radio Access Network (UE/RAN) for measurements.

UE-RAN positioning can also use Open Mobile Alliance Secure User Plane Location (OMA SUPL) with a SUPL Location Platform (SLP). RAN/UE traces via Trace Collection Entity (TCE) can also be processed by radio applications (rApps) to estimate device positions and correlate them with network events for Operations, Administration and Maintenance (OAM) and Key Performance Indicator (KPI) analysis.

There exist three key APIs for exposure: OMA Mobile Location Protocol (MLP), CAMARA and 3GPP Ngmlc.

Current solutions available in commercial networks include:

1. 3GPP passive data for network-based positioning, based on data available from non-positioning mechanisms such as mobility management.
2. 3GPP device agnostic signaling and RAN measurements for network-based positioning, based on activated signaling in RAN to enable positioning.
3. 3GPP RAT-dependent positioning, based on Rel 16 signals, measurements and procedures.
4. 3GPP RAT-independent positioning, encompassing device-based GNSS positioning with optional device reporting.

From a device capability support, two categories stand out:

A. Device agnostic category, only supporting fundamental mobility measurements and procedures – supportive of positioning mechanisms 1 and 2.

B. Device dependent category, supporting positioning mechanisms 3 and 4.

For Local Dedicated Networks, the dominating positioning mechanisms would be 2, supported by device type A). For public networks, positioning mechanisms 2 and 4 will be important.

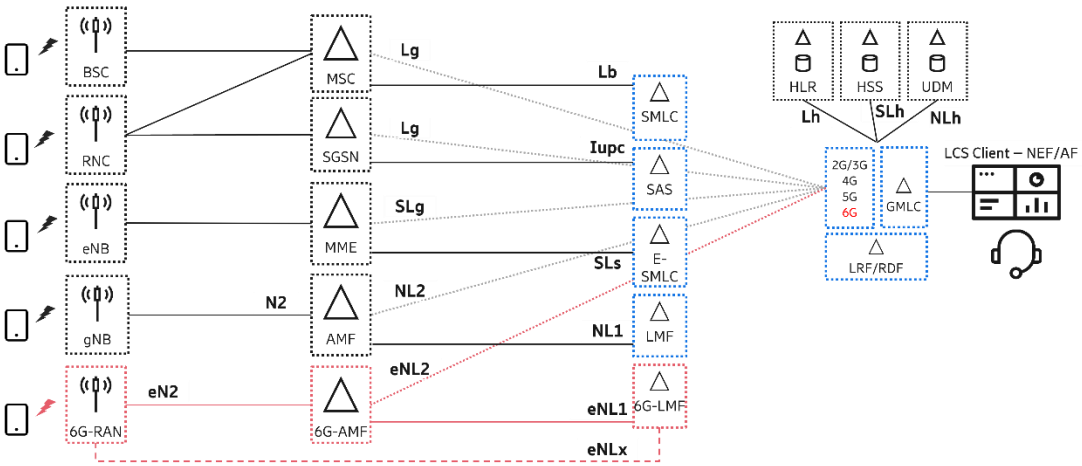


Figure 9 3GPP Positioning architecture with 6G Extension

The 5G positioning architecture is very fitting in many aspects, but some improvements can be considered, such as optimized support for periodic measurements and due to increased data transport needs.

Positioning will continue to be important in 6G not only for commercial but more so for regulatory requirements. AI-based positioning enhancements can be expected as can further additions of time referencing and the ability by the network to support adequate location exposure APIs and API transformations.

4.8 Integrated Sensing And Communication (ISAC)

Wireless sensing is a technology where a mobile system can acquire information about characteristics of an environment and/or objects within it. A key advantage is observing the environment without requiring objects to participate, without requiring objects to participate, i.e., objects observed do not have to be or include a user equipment (UE).

When sensing is integrated into a communication system such as a mobile system, it is called integrated sensing and communication (ISAC). Sensing in a mobile system uses radio frequency (RF) to detect objects and determine characteristics such as distance (range), angle, velocity and shape. 3GPP is studying how to include initial sensing capabilities for limited use cases in 5G Advanced, while 6G plans envision sensing for further use cases. Examples of possible market opportunities include ensuring public safety (for example, detecting and tracking drones), monitoring automated guided vehicles and preventing collisions, improving communication (for example, tracking signal-blocking objects), etc.

To address these needs, Ericsson has developed an end-to-end (E2E) architecture for 6G sensing.

In our E2E sensing architecture, shown in Figure 10, the 6G RAN is responsible for carrying out the sensing measurements required to fulfil the requirements on the sensing service, as requested by the sensing client. The workflow for the sensing service, i.e., the steps and actions to take to deliver the expected results to the sensing client, as well as

applicable parts of sensing processing are handled by the core functionalities for sensing or simply sensing core functionalities. The sensing core functionalities must handle policies related to the external sensing request (allowed areas, privacy policies and so on) when initiating a sensing measurement throughout workflow handling.

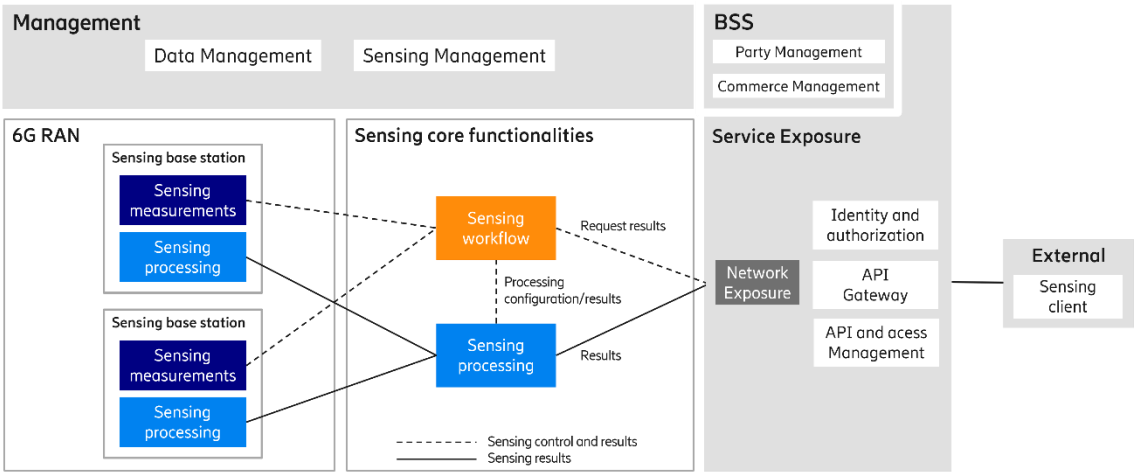


Figure 10 End-to-end sensing architecture including managements exposure and business support systems

Sensing use cases show potential to deliver benefits for consumers, enterprises and governments (especially public safety organizations).

Standardized mechanisms will however be needed for the network and the application to set up the sensing process and use an API to provide the results to the application (Exposure). Sensing may also be combined with other services such as positioning, SIM (subscriber identity module) density and other types of information for the benefit of applications that require comprehensive situational awareness.

Figure 11 illustrates the main types of capabilities under discussion for sensing, including the ability to detect objects and sense properties such as weather phenomena. The most obvious capability is detecting moving objects (A) through radio-based sensing mechanisms such as Doppler analysis or similar techniques. This can be valuable, e.g., in the protection of no-fly zones such as airports or stadiums.

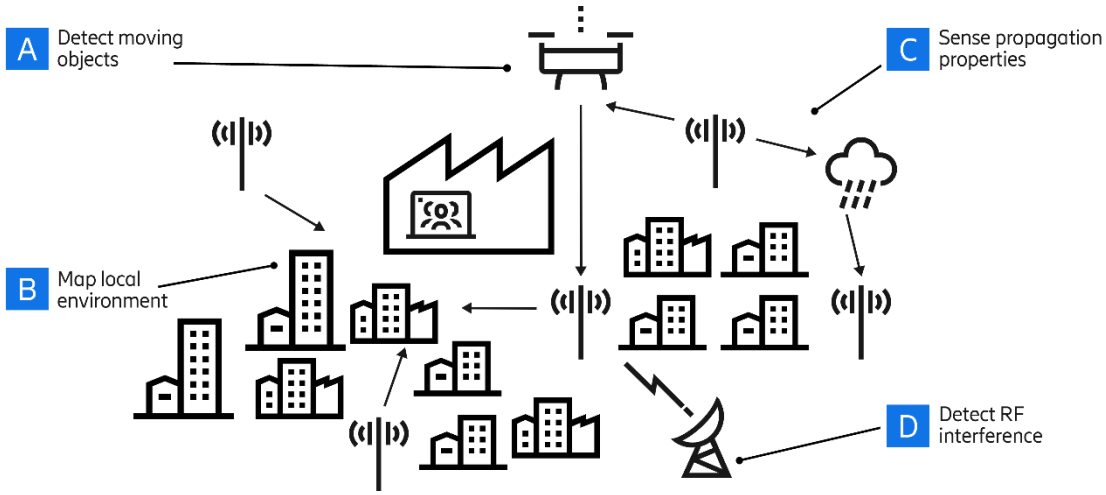


Figure 11 Sensing capability categories

To ensure optimal radio resource utilization, sensing measurements should be both executed and handled in RAN. An additional benefit of this approach is that it ensures the availability of RAN internal processing resources to handle the sensing measurements.

While 5G Advanced will introduce the first standardized sensing features, it is in 6G that integrated sensing and communication is likely to reach its full potential.

Related articles/additional reading:
Sensing 6G: Use cases and architecture

5 Network architecture domains

5.1 6G Network Architecture Direction – The 2030 perspective

The telecom industry and academia have been researching 6G wireless technology since more than five years and initial 6G standardization work in 3GPP has started.

6G will act as a foundation for a cyber-physical continuum, merging the digital and physical worlds, supporting extreme usage scenarios and massive scaling, with demanding applications across industries, society, and consumer domains.

Key capabilities such as ultra-low latency, high precision positioning, sensing, etc. will enable advanced industrial and societal applications while merged reality, digital twins, and collaborative robotics, will be driving innovation in mission-critical and consumer services.

Obviously 6G will be AI-native and Agentic AI in place from the start. This will be part of introducing automation, based on intent for self-optimization with a smooth and spectrum-efficient integration leveraging multi-RAT spectrum sharing and evolving 5G core components.

Regarding the introduction of 6G we should remember that we are around mid-way through 5G and a top priority for operators today is how to monetize 5G SA and 5G Advanced capabilities to improve top line revenue reusing and expanding on the evolution of 5G.

At the same time, business efficiency includes the automation of processes handling of the full lifecycle of customers, partners, suppliers, products, orders etc., from both commercial and operational aspects. Automation will complement and eventually replace manual work to develop, deploy, manage and optimize the mobile network and includes intent-based management.

Above evolution includes adoption of Artificial Intelligence and Machine Learning (AI/ML) technology including generative AI as well as AI agents which are needed to simplify network operation and optimization.

Data is recognized as a crucial asset and is already a fundamental part of the CSP's network.

It is expected that future mobile networks continuously benefit from the continuous breakthroughs in the areas of cloud native adoption, automation of network operations and AI/ML.

A central trend is about network architecture evolution itself. To support the 6G vision and requirements, it is beneficial to standardize a 6G architecture that allows for a smooth introduction of 6G capabilities into future public and private networks.

The top-level view for a 6G network starts with the Global Network Architecture, shown in Figure 1 above, representing a transformational view in how networks must be built, operated, and opened up for innovation. Instead of dedicated, well-defined, and vertically integrated nodes connected in a static network setup, the networks are evolving towards a more dynamically adaptable architecture where Network Functions (NF) and applications are running where and when they are needed to optimize performance, cost, and business agility.

6G should be used to leverage and expand the usage of the network as an innovation platform, realizing more of the platform capabilities in the real networks in parallel with the evolving ecosystem.

Separation through horizontalization between HW, cloud, transport, data pipelines, network applications, management, and monetization will consequently make interfaces supporting this horizontalization more important for multi-vendor deployment.

In this transition, correct modularization of network functionality is crucial. An open radio interface, open RAN-CN interface and global roaming specified in 3GPP are the remaining basis to support the global scale representing a strong ecosystem. Figure 12, below shows the 6G Network Domain Architecture.

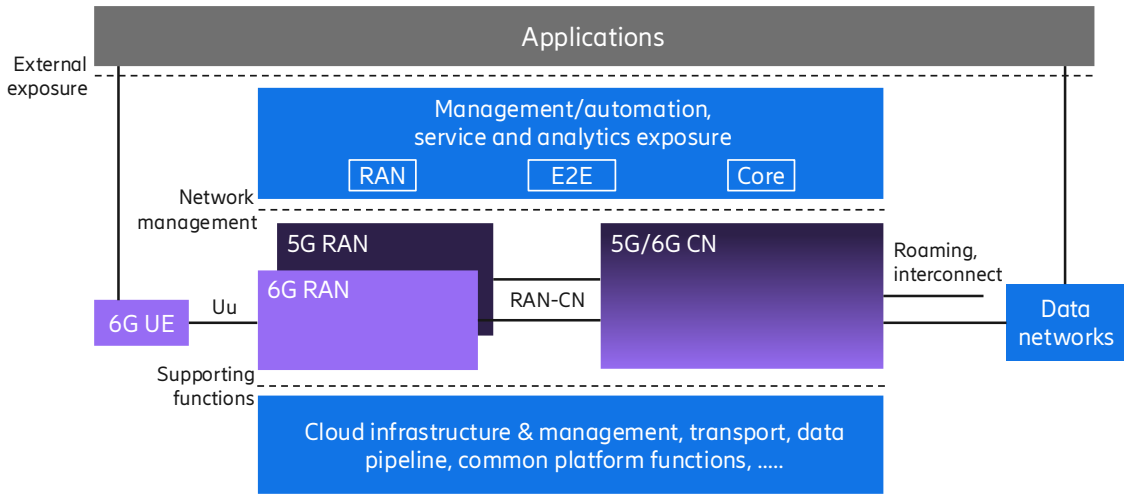


Figure 12 Key Open Interfaces in the high level 6G domain architecture

The following categories are examples of such open interfaces are shown:

- **Uu,** a new stand-alone 6G radio interface operating in new / existing frequency bands, efficient spectrum co-existence with legacy RATs
- **Roaming, interconnect** for control and user plane evolved from 5G
- **RAN-CN** interface, as an example of a key network-internal standardized interfaces incl. also RAN internal (LLS), RAN-RAN, CN internal and CN-CN.

In addition to the interfaces shown in Figure 12 there will be interface(s) enabling vendor-specific exposure for data sharing, CI/CD, SW delivery pipelines, managed service, etc.

Related articles/additional reading:
6G network architecture – a proposal for early alignment
Co-creating a cyber-physical world

5.2 Radio Access Network (RAN)

RAN's main purpose is to transfer end-user application data, that may have different patterns and needs, between UE and CN, by using provided resources such as spectrum, sites, hardware and transport. Long-term, an automated self-optimizing RAN that optimizes RAN performance in a geographic area (rather than per node/board/cell), steered by intents, (not excluding other solutions) given a set of provided resources can be envisaged.

The performance of a RAN covers several dimensions that increase end-user experience or decrease the CSP's TCO. Such examples include data rates, latency, capacity, coverage, energy consumption, TCO, ease of use, NRAR, etc.

The importance of above dimensions varies between different contexts and for example, the desired performance in different dimensions will depend on service categories (service differentiation) and end-user groups (subscriber group & slice differentiation), while NRAR will be imperative for Mission Critical Networks.

The long-term evolution of a self-optimizing RAN means that the management interface towards RAN has evolved to describe *what* RAN shall optimize for (intents for the RAN service) rather than *how* RAN shall operate (setting thousands of parameters related to traditional FCAPS management).

RAN automation is integrated in all levels of the RAN functional domain and includes high-level RAN automation (rApps and SMO platform capabilities) as well as automation in the RAN NF's and RAN will be AI native.

The current best thinking of a RAN standard target architecture is shown in Figure 13. It illustrates that the 6G RAN consisting of 6GNB's, which can internally be connected with an LLS interface (e.g. an evolution of the O-RAN Open Fronthaul interface (OFH)), into a RAN Area function (RANAF) and one or more Radio Units (RU).

High-level automation functions deployed in SMO can manage these RAN network functions via standardized logical interfaces for NF management (based on evolution of O1, A1, R1, m-plane).

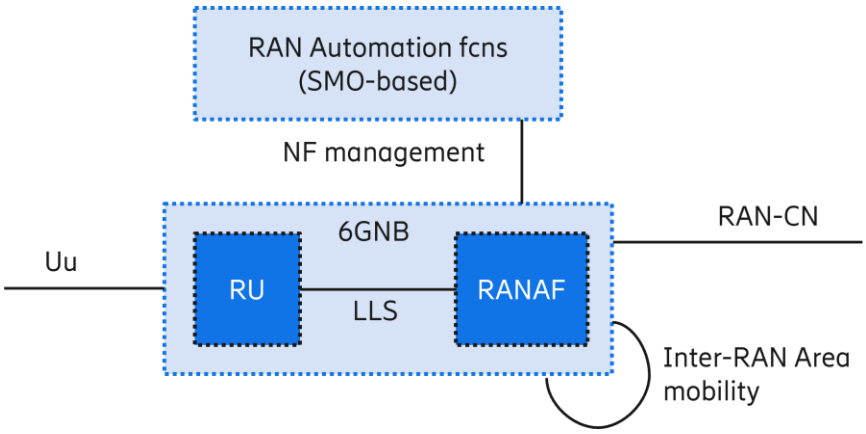


Figure 13 6G RAN standard target architecture

Ericsson is embracing all RAN-internal interfaces above as potential Multivendor interfaces, where we need to comply and be prepared to do interoperability testing, at least with basic/limited performance/features. However, it is expected that the main deployment model for multi-vendor RAN will continue to be by RAN vendor geographic areas, using interoperable Inter-RAN Area Mobility i/f.

In 6G the lower-layer-split (LLS)/O-RAN Open Fronthaul (OFH) interface between the RANAF and the RU should be standardized and implemented as a multi-vendor interface. Currently an (typically single-vendor-) LLS is used in practice in all RAN products and deployments thereof.

6G should be a new stand-alone RAT, meaning 6G capable UEs will connect directly to 6G capable base-stations when in coverage. Simultaneous 5G and 6G connectivity (similar to the NSA NR) should not be standardized since it will delay stand-alone 6G deployments, increase the overall development effort (maintain several tracks) and limit the performance potential of the 6G.

The 6G radio standard should support optimized spectrum sharing between 6G and 5G, and non-optimized spectrum sharing between 6G and LTE (incl. NB-IoT), aka Multi-RAT Spectrum Sharing (MRSS). This will quick support and efficient single vendor deployment of 6G in existing 5G bands maximizing spectrum, power and HW usage.

The theoretical overhead of 5G/6G MRSS is small enough to expect 5G/6G MRSS deployments to be very common.

To optimize RAN performance across spectrum and geography, efficient multi-dimension RAN coordination, Figure 14, can be done by using different resources across resource instances. It will be important to handle multiple dimensions of radio resource coordination, including, e.g. :

- Carriers / spectrum (e.g. Carrier Aggregation, and selection of serving sector carrier, UL/DL carrier decoupling)
- Transmission/Reception Points (e.g. Interference management, UL COMP and other multi-TRP / D-MIMO features)
- Inter-RAT spectrum sharing (e.g. 5G-6G spectrum sharing in the same sector carrier)
- Multi-band power pooling (how multiple bands supported by one Radio Unit can share one transmit power resource)
- Agile PIM avoidance (i.e. schedule to avoid PIM and then use full power without backoff)

In addition, there are non-radio resources to be managed, such as transport and processing HW resources. All these resources should as far as possible be treated as pools of resources which can be flexibly used depending on current needs.

Much of this RAN coordination also needs to happen fast enough to match the bursty nature of real traffic patterns.

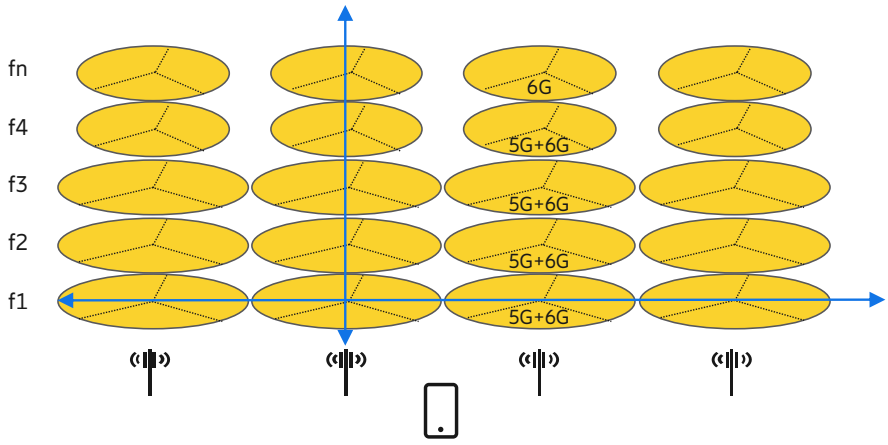


Figure 14 Multi-dimensional RAN coordination, across spectrum (e.g. CA) and antenna sites

With massive use of 5G-6G spectrum sharing, enabling use of legacy spectrum for 6G, there is a need for very fast and efficient sharing between the RATs within a sector carrier. This will be a challenge to be solved by RAN vendors and will likely require standards on the Uu to provide enablers for this. Also, other parts of the specifications should take height for that a given sector carrier can be shared between a 5G cell and a 6G cell.

Related articles/additional reading:
Advanced connectivity solutions for time-critical communication traffic in 5G radio access networks
Energy performance of 6G Radio Access Networks: A once in a decade opportunity
6G spectrum - future mobile life - Ericsson

5.3 Core Network (CN)

The core networks for initial 6G deployments need to address a range of different aspects to be relevant in 2030 and beyond. A first aspect is to efficiently support multiple access network generations. This applies in particular to 5G SA to provide an optimized mobility between 5G and 6G. This is instrumental for a smooth introduction of 6G.

Therefore it makes sense that 3GPP will use the evolved 5GC as baseline when defining the 6G CN support for a 6G radio network, in contrast to defining a new 6G core network. There may be new features in Core to support the 6G RAN and 6G UEs, but these should be considered from the angle of having limited impact on the overall architecture.

There are several reasons why 6G CN should evolve from 5GC such as the need to support LTE and NR interworking beyond 2030 or simply leveraging industry's investments already made in 5GC (3GPP, vendors and CSPs) as well as continued monetization leveraging already deployed 5GC enabled services based on e.g., the granular Quality of Service framework, network slicing or time-sensitive communication capabilities.

To address these existing use cases supported by 5GS, the resulting 6G CN is outlined in Figure 15. This basic 6G support impacts all NFs in 5GC to a lesser or larger extent, e.g.

- AMF handling both 6G-RAN and NG-RAN
- SMF and UPF single anchor point with 6G enhancements
- UDM handling 5G & 6G user data
- PCF handling 5G & 6G policy data

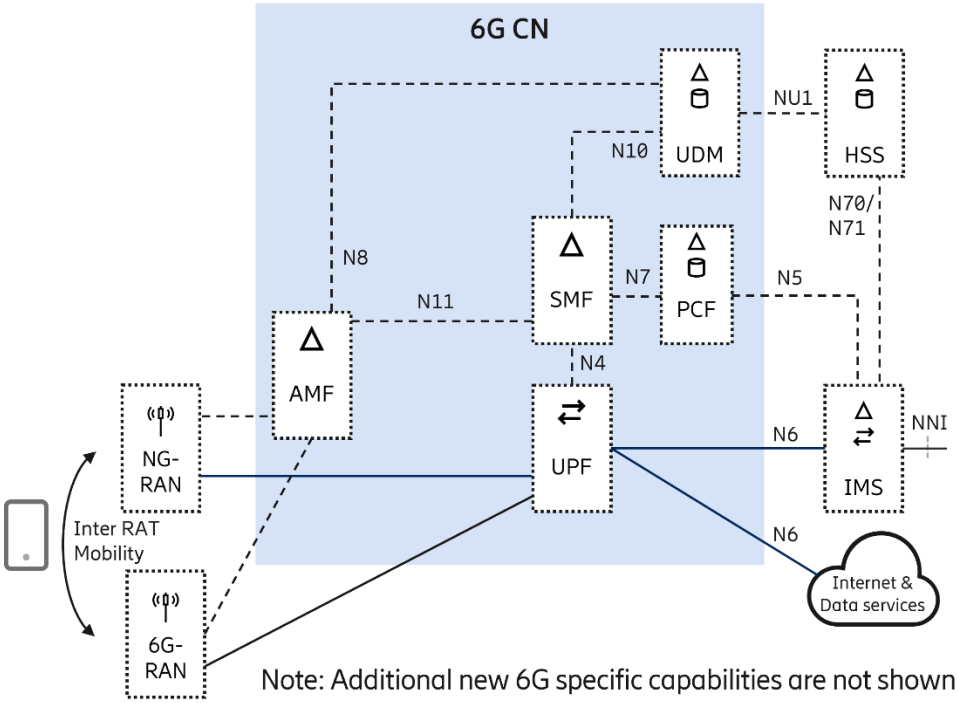


Figure 15 6G Core Network (CN) – based on an evolved 5GC to support 6G RAN

Furthermore, to enable new business opportunities the 6G CN will include additional functionality, including new NFs for new 6G specific capabilities and services, which should be defined with the above architecture as a baseline. Following are some examples of such new capabilities under investigation:

Integrated Sensing and Communication (ISAC), described in chapter Integrated Sensing And Communication (ISAC). Anticipated to work in coexistence and leverage other existing NFs like NRF for registration and discovery, and NEF for exposure.

AI will be a central part of the evolution to 6G as outlined in chapter AI in the Network architecture and used as a realization technology inside NFs as well as for optimization tools near the core network.

A **distributed data infrastructure**, with a network wide framework for data collection and management, required to support many use cases, like AI, see chapter Data Handling

Massive-IoT will be part of the first release of 6G, which gives an opportunity to correct some of the drawbacks in the 4G/5G standardization of M-IoT, such as adopting traffic models to actual usage, etc.

6G NTN will further evolve the deployment options and integration between TN and NTN networks. From a core network perspective there are multiple aspects to consider depending on the chosen deployment model. The main integration model between TN (MNO) and NTN (SNO) would be to reuse the roaming interfaces.

Energy efficiency in core network deployments is another area requiring attention, with potential power savings through implementation architecture and cloud infrastructure.

Related articles/additional reading:
Telecom evolution toward 6G

5.4 Communication services and Immersive Communication

The telco industry view is that IMS will serve as communication engine also in 6G, for 3GPP and non-3GPP access (e.g., Wi-Fi, NTN), providing, at a minimum regulated telephony, emergency services and SMS in wide area networks enabling support by the 4G and 5G IMS solution, with minimal changes. The same position applies to SMS, both for SMS over NAS and IP.

This allows for reuse of the investment and the user experience of these services, simplifying the interworking with 5G and 4G. To protect earlier investments, the 5GS roaming should be reused, with minimal changes, to mitigate the roaming challenge of supporting roaming to 2G/3G still in use in the 2030-time frame.

For a smooth 6G introduction, the CSP should migrate to an IMS based voice solution prior to launching 6G. Only minimal changes to IMS are foreseen, e.g., to address 6G location to enable IMS in 6G system (6G enabled IMS). This will be an evolution of the existing voice in 5GS architecture. It is further recommended to only evolve the SBI interfaces of IMS to avoid evolution of Diameter interfaces for 6G.

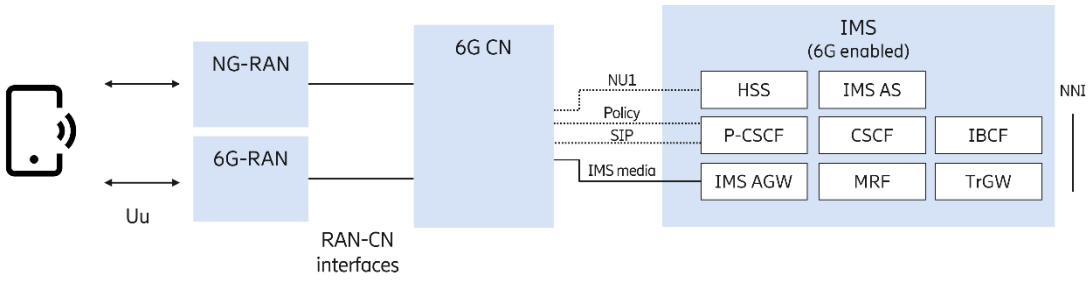


Figure 16 Functional target architecture – Voice in 6G architecture

The evolution of communication services to immersive communication in 6G will span different technologies like; enabling web experiences in the context of a telephone call (IMS DC), Immersive visual and audio experiences (XR), introducing AI agents/assistants, in the context of telephony, etc.

Technology and architecture for realization of new conversational use-cases is an ongoing industry discussion for 5G already.

Ericsson is exploring and evolving our architecture to embrace the use of AI in service execution. Architecture needs to support interaction with agents, between agents and with supporting AI capabilities like Language models, multi-modal models and tools.

Our aim is to evolve a single IMS architecture to support multiple use-case categories, where application logic can reside fully under CSP control, but also open for interaction with 3rd party, including Exposure/APIs and Media.

High level architecture to enable telephony enrichment use-cases is depicted below, Figure 17 including the examples of web, XR and AI.

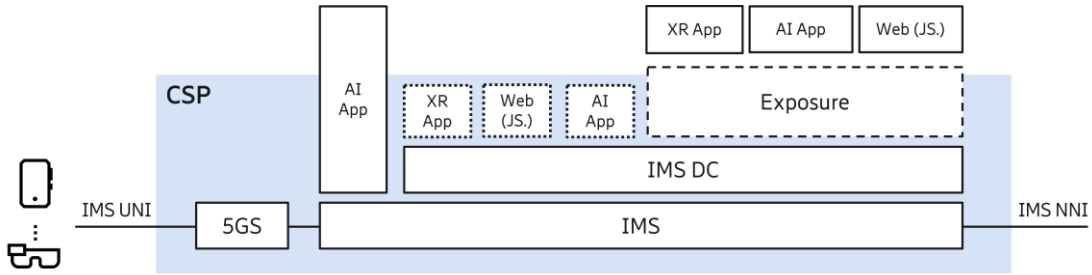


Figure 17 One IMS architecture for immersive use-cases

5.5 Data Handling

Data handling is a cross-cutting concern covering multiple domains of data sources and data of varying characteristics. Data management capabilities will become a fundamental part of the CSP's architecture as it is key to several processes.

Over the years multiple independent data pipelines for specific purposes have been created by CSPs, creating an unnecessarily rigid and difficult situation, resulting in data handling inefficiency and challenging governance of data.

The benefits of a common consolidated data ingestion for CSPs include significant OPEX and CAPEX cuts and possibly increased average revenue per user (ARPU).

To create this consolidation, Data Meshes can be created in a CSP domain (DMF) using EDCA functionality incorporated into our products and in the Ericsson domain, the Ericsson Federated Data Lake (EFDL). See Data Ingestion Architecture below in Figure 18.

A CSP can leverage consolidation of data in several ways one of which is using it to enable Network Automation. The data could also be exposed through Network APIs or playing a central role in AI/ML model training acting as a pipeline for refinement into a data pipeline for AI/ML created from data ingested using EDCA functionality.

Several data collector instances may exist in a network and be part of multiple functions in the network working in a federated approach – e.g. being part of the SMO as one data set of data islands.

In a customer network, the data mesh concept can be used to tie data islands in several Ericsson products into one data mesh which may also be connected to a corresponding data mesh in the EFDL as shown in the Data Ingestion Architecture below.

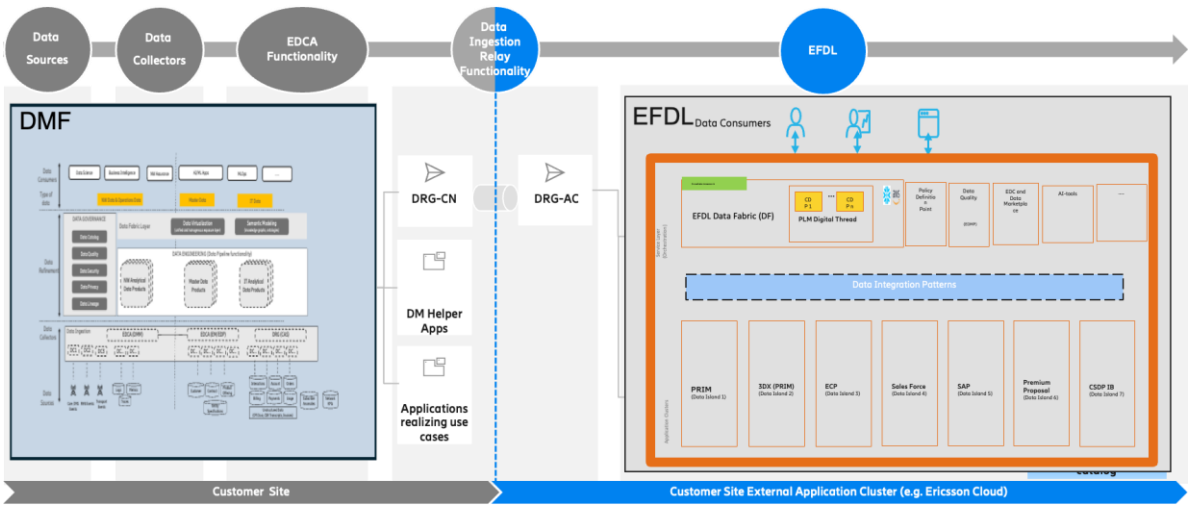


Figure 18 Data Ingestion Architecture

By creating a data mesh, data collected in a data mesh island is mirrored to other data mesh islands if the same data is required by these. From an application point of view, data in all data mesh islands can be discovered, consumed and leveraged, provided that the application has the required access- and authorization level.

Applying a data fabric in the customer network and on the Ericsson side, turns the two data meshes into a universal data mesh. The same principle can be applied in the reverse direction and allow Ericsson applications in a customer network to access data in one of the Ericsson application clusters.

The Data Fabric and EFDL will decrease the extra load on the transport network deployment as manipulations to the data can be made without moving data. This is also referred to as a "no data copy solution" to separate it from centralized solutions.

A CSP can leverage the EFDL for a number of other solutions to improve their network business in areas like support and serviceability, extended AI/ML training, evolved LCM (DevOps, MLOps, etc.).

Another advantage is that a CSP can leverage access to data from the EFDL to draw on analytics insights, reports, data for AI/ML training. This achieved via a Data Fabric which applies the necessary access policies, contracts, legislation adherences, etc.

As a reminder some of the principles for Data Handling are listed below:

- Collect data once, use many times
- Manage Data in a federated approach with multiple Data islands
- Data should be used in transparent, compliant and ethical ways
- Data Quality must be ensured during the data lifecycle
- Data access needs to be authorized to enable legal and ethical use.

6 Network architecture examples

6.1 Network Deployment Cases – Private Networks / Enterprise

Enterprises depend on business processes to govern their operations. These span across internal structures like organizations, departments of the enterprise, involving various assets, people, and operational systems (primarily IT – Information Technology, and OT – Operational Technology) of the enterprise.

An Enterprise Architecture typically consists of a business domain and a technology domain. Here we will focus on the technology domain. Figure 19 shows a combined view of the technology domain of the Enterprise Architecture and the 5G/6G Network Architecture, including a simplified representation of the main interfaces between the two, leaving out certain details for simplicity reasons.

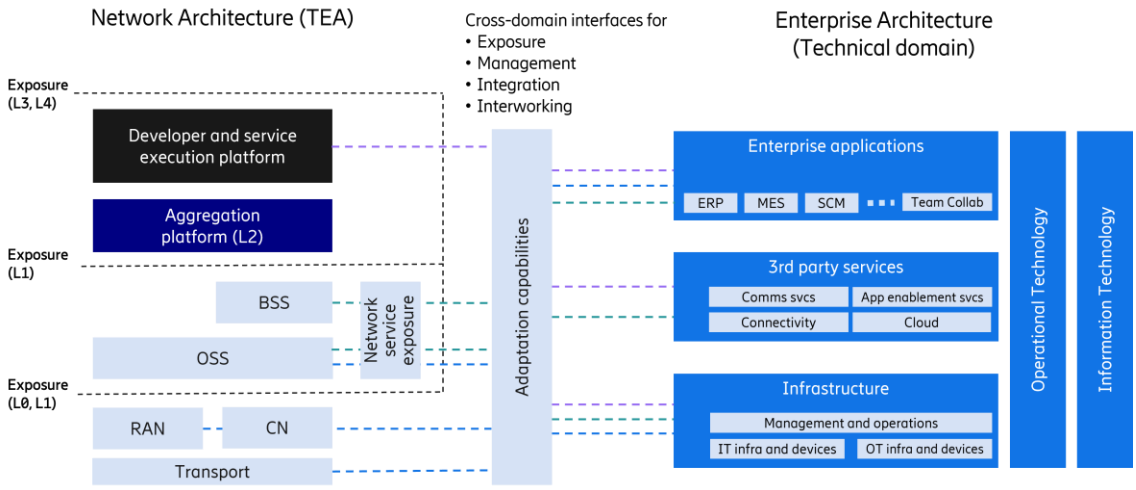


Figure 19 Summary of main interfaces towards an overall enterprise architecture

These interfaces can be summarized in three main categories:

- Networking, connectivity, and associated management (blue)
- Overall management functionality (green)
- Consumption of communication services, e.g. voice, video, messaging (purple)

3GPP based telecom networks are highly suitable as the base architecture to support the technical side of an Enterprise architecture for many different scenarios, enabling multiple use cases for several customer types.

Several deployment models are possible for cellular networks targeting enterprise customers and best suitability depends on the needs and requirements of the enterprise in terms of e.g. data privacy, ownership of network resources, level of operational control, geographical reach, etc. In some cases, a combination of deployment types may be relevant for a particular enterprise.

Network deployments for enterprise can largely be classified in four main models based on the amount of dedicated versus shared network resources as shown below in Figure 20.

Derived from widely used 5G-ACIA NPN deployment models. Only a broad classification, not exhaustive. Many combinations and variations are possible.

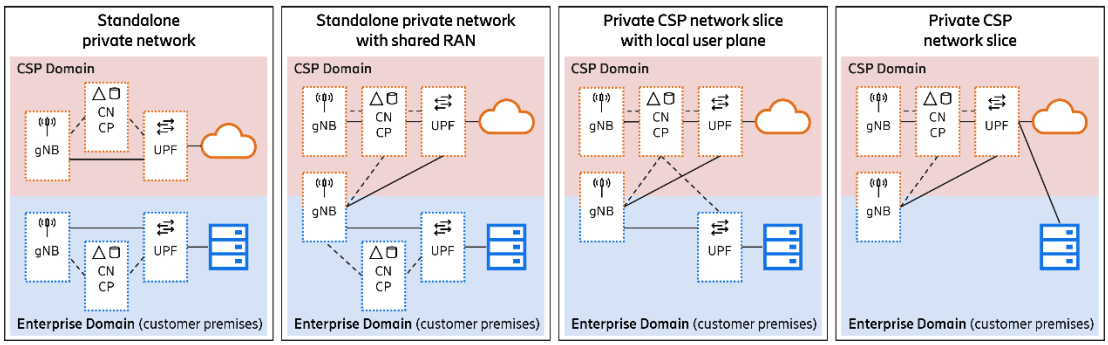


Figure 20 Deployment models for enterprise

In a standalone private network deployment, network resources and equipment are dedicated to the sole use of the enterprise, with sole authority to determine which users and devices can access the network. This type of deployment is typically confined to a localized area (e.g. within a building, port, etc.). The standalone private network with shared RAN is a variant where the radio infrastructure is shared (MOCN or MORAN) between the enterprise and CSP.

The two other deployment models are variants of network slicing provided by a CSP, sometimes referred to as virtual private mobile networks, which differ in the amount of dedicated network resources deployed at the customer premises. The variant with local user plane ensures that the data stays within the premises may be more suitable for time-critical applications. Both variants will typically have radio equipment deployed on-premises, however it may or may not be dedicated (i.e. restricted) to exclusive use by the enterprise.

A further variant of network slicing, not shown above, is the so-called infrastructure-light approach where the only dedicated equipment in the enterprise is the managed endpoints (laptops, smartphones, WWAN gateways) that connect directly to the CSP domain via a private slice.

Seamless operation when moving between different types of deployment is expected to become more important in the future, e.g., logistics-related use-cases in mining, factories, etc, where vehicles may move between a private network and a slice of a public network. This applies both to exposure- and communication interfaces.

The network should provide a unified experience to applications and users regardless of the type of deployment. This is very important from a developer perspective to facilitate portability of applications and use cases and to accelerate their development, or simply put, to enable scale.

Related articles/additional reading:
[Reducing handover interruption with L1/L2 Triggered Mobility]

6.2 Mission Critical Networks (MCN)

This chapter describes four MC segments; Public Safety, Defense, Utilities and Rail and three technology areas; ISAC, Non-Terrestrial Networks and Digital Airspace which are applicable both to MC segments and non-MC business. This architecture document is focused on architecture needs to meet MC driven demands.

What characterizes all MC segments is the criticality of communication service as they need to be able to rely on the networks and services provided especially in challenging scenarios with partial loss of e.g. infrastructure or power. If the network fails it is not only a matter of major business impact, but it may even be a matter of life and death.

There are commonalities across the MC segments as well as unique aspects for each segment, defined in below sub chapters.

Figure 21 shows some essential aspects, of several, which are applicable to all MC segments:

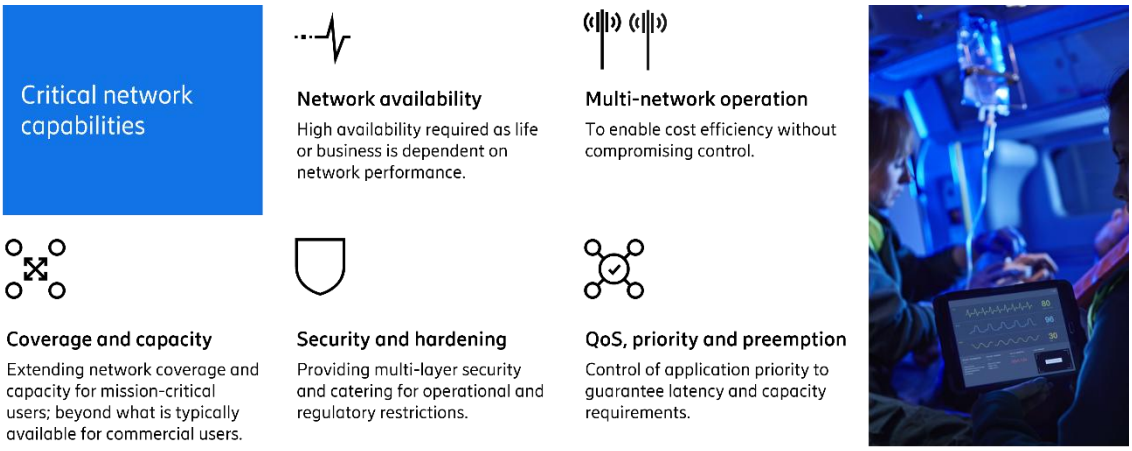


Figure 21 Essential aspects of Mission Critical Networks

There is a strong need for Network Reliability, Availability and Resilience (NRAR) for the end-to-end connectivity service, as well as non-connectivity services like ISAC, chapter Integrated Sensing And Communication (ISAC) and positioning, chapter Positioning .

Furthermore there are expectations of simplified deployment, maybe leveraging the self-x of Autonomous Networks and operations of small and flexible Mission Critical Networks especially in Defense Networks.

Security in MCN largely follows product security requirements MBB networks but with more strict demands on levels of maturity and compliance. Examples of security areas to consider are Post Quantum Cryptography, Zero Trust Architecture described in the chapter Security

Needless to say, AI will play a central role also in these network for example in incident response, monitoring and alerting, etc.

6.2.1 Public safety

The Global migration from LMR (e.g., Tetra) to 3GPP based networks for nationwide Public Safety is steady but slow. The deployment models vary by country, driven by spectrum availability and willingness to reuse CSP infrastructure. We notice an emerging trend towards cooperation between governments and incumbent CSPs.

Early adopters drive a "digital office in the field" use case with additional improved situational awareness through real-time video, location, and sensor data.

Most Public Safety spectrum/use today is LTE and as the ecosystem matures a migration path towards 5G will happen.

Some operators offer mission-critical slices on 5G SA; however these are often premium offerings rather than full 3GPP-standard MC services.

For 5G resilience, features such as NTN, Wireless Access Backhaul (WAB), UE-NW relay, and disaster roaming can be combined. 6G architecture should further strengthen resilience for both society-critical and mission-critical needs.

Different deployment options are available to cater for different situations depending on several conditions like: Availability of spectrum, Spectrum Regulations, Operational model chosen by the government, etc as depicted in Figure 22.

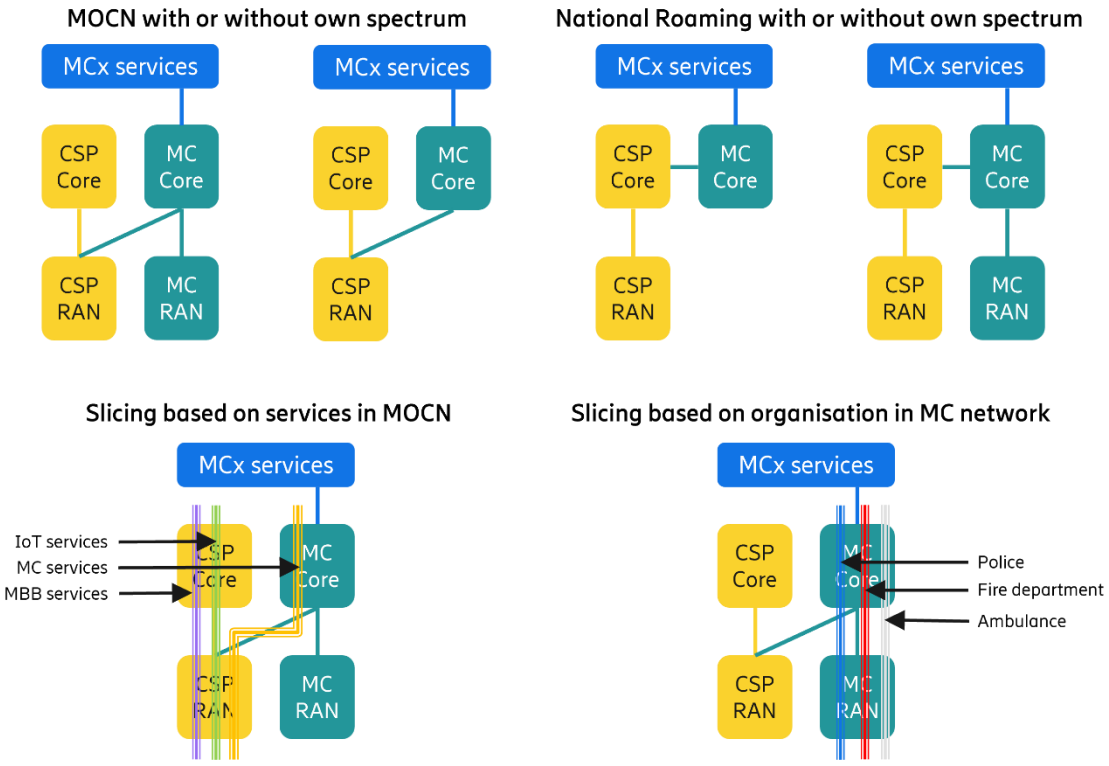


Figure 22 Deployment options, with or without slicing, non-exclusive

If there is no dedicated spectrum available, then RAN-sharing with a CSP is the only

option. The operational model will dictate if MOCN, National roaming, slicing etc is possible and how.

There is a growing need for Cross-border collaboration between Public Safety organisations in different countries, e.g. wildfires or earthquakes or major events requiring support from different countries in a concentrated area.

6.2.2 Defense

Geopolitical volatility is driving increased focus on strengthening defense and resilience. Modern warfare is turning into an increasing use of technology (drones, sensors, satellites), making the battlefield more transparent with increasing detection/elimination risks.

Decision-making is becoming faster and more data/AI-driven, which increases the demand for high-throughput communications close to battle zones.

Above developments and experience from recent conflicts emphasize the need for severe improvement to the existing military communications.

Lessons from the Russia–Ukraine war emphasize digital transformation, use of commercial technologies, and federated communications for massive information exchange.

Armed forces use the PACE (Primary, Alternate, Contingent, Emergency) model; commercial cellular systems are expected to complement military comms as primary or alternate options depending on mission. This has led to exploration of partnerships with public MNOs jointly with own private networks. Military organizations like NATO are standardizing for cellular use to interoperate with legacy systems and enable Federated Mission Networking (FMN).

Needless to say ISAC will be an important technology, see Integrated Sensing And Communication (ISAC)

6.2.3 Utilities

Digitalization, decentralization, and demands for security, resilience, and flexibility elevate Mission-Critical Networks (MCNs) to the role of the grid's operational backbone.

Current constraints include siloed OT/ICT stacks that inhibit deterministic control, real-time analytics, and coordinated automation. Utilities require nationwide reach with deterministic local performance (substation-level latency, reliability).

The resulting architecture converts a network-centric MCN into an integrated system of connectivity, computation, and control optimized for grid operations that unifies data, control, and decisioning across the energy value chain, an intelligent Digital Nervous System, Figure 23.

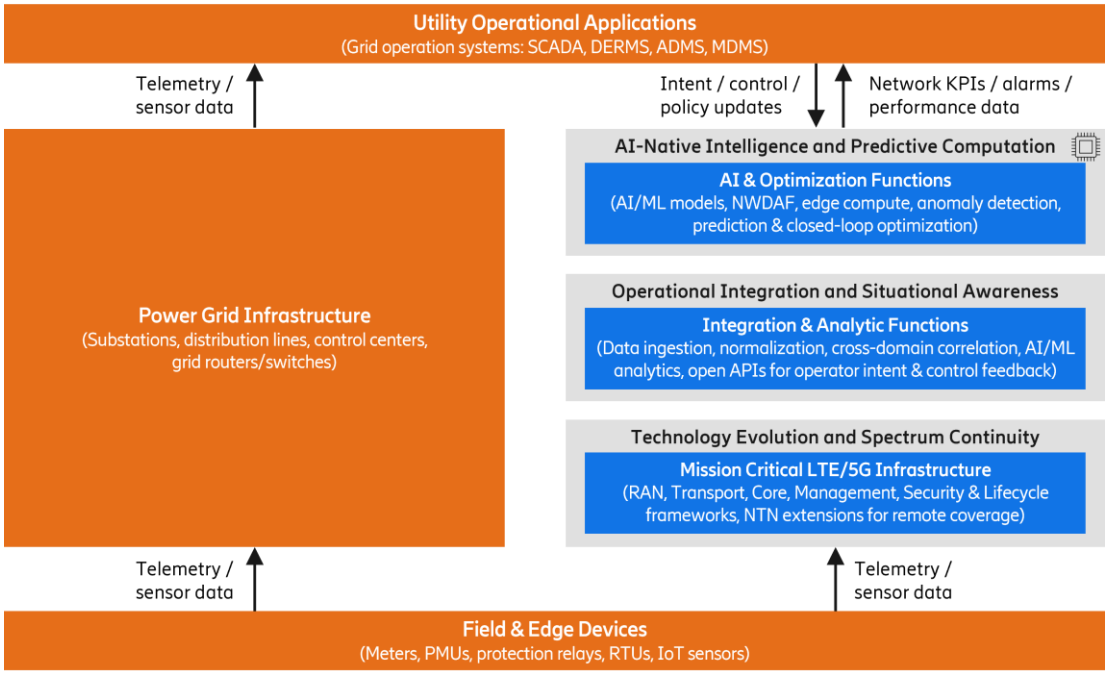


Figure 23 MCN as the Digital Nervous System of the Utility Grid

6.2.4 Rail

FRMCS architecture includes support for safety-critical train control, fully redundant deployments including baseband redundancy, precise positioning and supports gigabit passenger connectivity.

FRMCS standardization and validation is planned for completion in Q4 2027.

Cross-Border Operations will demand seamless and instantaneous transition of train control across borders requiring local breakout Roaming for direct connection to visited core systems while retaining select home-routed applications.

7 Abbreviations and Definitions

3GPP	3rd Generation Partnership Project
5GC	5G Core
5GS	5G System
AI	Artificial Intelligence
CAPEX	CAPital EXpenditures
CN	Core Network
CPaaS	Communication Platform as a Service
CSP	Communication Service Provider
DL	Downlink
EDCA	Extensible Data Collection Architecture
EFDL	Ericsson Federated Data Lake
ETSI	European Telecommunications Standards Institute
FRMCS	Future Rail Mission Communication System

GDPR	General Data Protection Regulation
IMF	Intent Management Function
IMS	IP Multimedia Subsystem
ISAC	Integrated Sensing And Communication
LCM	Life Cycle Management
MBB	Mobile Broadband
MRSS	Multi-RAT Spectrum Sharing
ML	Machine Learning
NF	Network Function
NI-QoS	Network Initiated Quality of Service
NIST	National Institute of Standards and Technology
NRAR	Network Reliability, Availability and Resilience
NTN	Non-Terrestrial Network
OPEX	OPeration EXpenditures
ORAN	Open Radio Access Network
O-RAN	Open RAN
OSS	Operation Support System
OTT	Over The Top
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
SDG	Sustainable Development Goals
SLA	Service Level Agreement
SMO	Service Management and Orchestration
SOM	Security Orchestration and Management
SW	Software
TC	Traffic Classification
TCC	Time Critical Communication
TCO	Total Cost of Ownership
TMF	Telecom Management Forum
TN	Transport Network
TTM	Time To Market
UE	User Equipment
UL	Uplink
UP	User Plane
URSP	UE Route Selection Policy
VoLTE	Voice over LTE
VoNR	Voice over NR
ZTA	Zero Trust Architecture

Intent can be defined as a "formal specification of all expectations including requirements, goals and constraints given to a technical system". It states which goals to achieve rather than how to achieve them. Intent enables the creation of autonomous sub-systems rather than creating tightly coupled management workflows.

8 References

[1] https://www.itu.int/rec/R-REC-M/recommendation.asp?lang=en&parent=R-REC-M.2160

[2] https://www.ericsson.com/en/reports-and-papers/mobility-report/articles/high-performing-programmable-network-atandt

[3] NIST SP 800-207, Zero Trust Architecture, https://csrc.nist.gov/pubs/sp/800/207/final

[4] Introducing TM Forum's Autonomous Network Mission: The roadmap to true autonomy in telecom