Ericsson White Paper
GFTL-26:000204 Uen
January 2026

From data mess to AI-ready data mesh

Content

Introduction

Over the next decade, the telecommunications landscape will be shaped by the ability to unlock and utilize the full potential of data at an unprecedented scale, speed, and intelligence. As a trusted partner, we have observed how sprawling Hadoop architectures and on-premises platforms in the telecom sector can quickly become brittle as volumes expand and new use cases emerge.

In addition to the expansion of data volume, data fragmentation across domains, platforms, and schemas is an inevitable consequence. Autonomous network domain elements, service assurance engines, customer experience agents, and other similar systems produce domain-specific silos, which may delay the highly multi-agentic operations that communications service providers (CSPs) seek to achieve.

In the data management layer, the challenge is to manage the integration of high-volume data inflows from disparate sources, govern the movement and transformation of that data securely and with integrity, and bring that data to a state of readiness for a variety of consumers, most notably those who will apply various artificial intelligence (AI) techniques.

Hence, a strategic augmentation of CSP data management architecture is required—one that natively anticipates domain fragmentation by design, enforces seamless and secure exchange, and guarantees readiness for a multi-agent system from day one.

This document captures thought leadership and a reference architecture for a truly future-proof data management that scales elastically, supports seamless data integration, can be deployed in a hybrid manner, and is purpose-built for AI-native intelligence and autonomous network and operations.

Background context

This chapter captures the various industry, technological, and business trends that are influencing the nature of and expectations on data management layers within organizations.

Industry trends

The growth in data volumes continues to exceed predictions as networks continue to grow. Telecom operators have existing data management architectures, but the costs are high, as large datasets usually operate with unnecessary data transit and storage.

The current growing data demands of the telecom industry present two problems. Managing large volumes of data will continue to be important and unavoidable, but that is a well-known engineering issue. However, a more uncommon problem is the ability to curate data from diverse sources to ensure its efficient and appropriate use.

Technology trends

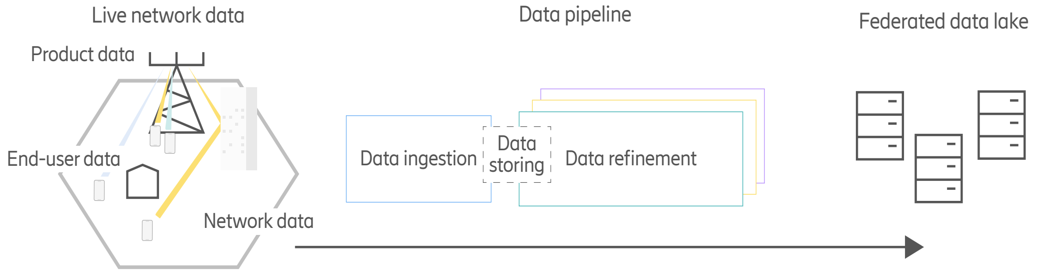
The most obvious driver in terms of technology trends is the increased use of AI, which includes agentic AI, robotics, and automation associated with AI. The use of AI demands higher data quality and quantity, which means better data governance, integrated datasets, and the articulation of relationships. AI also further motivates the need to make data more accessible, which is a significant driver for the establishment of a data fabric, enabling a unified approach to data access through integrated catalogs and other federated services.

Business trends

A significant proportion of businesses are investing heavily in automation and autonomous networks. By the end of 2026, approximately 80 percent of businesses will accelerate their automation efforts to streamline operations and maximize revenues, with most of them focusing on leveraging AI to make meaningful gains in these areas.

This will put extra emphasis on data, its timeliness, availability, and readiness for use in AI processes. As already mentioned, data availability and readiness are critical in the data management domain, which demands high standards in many other related aspects of data management.

Data management evolution



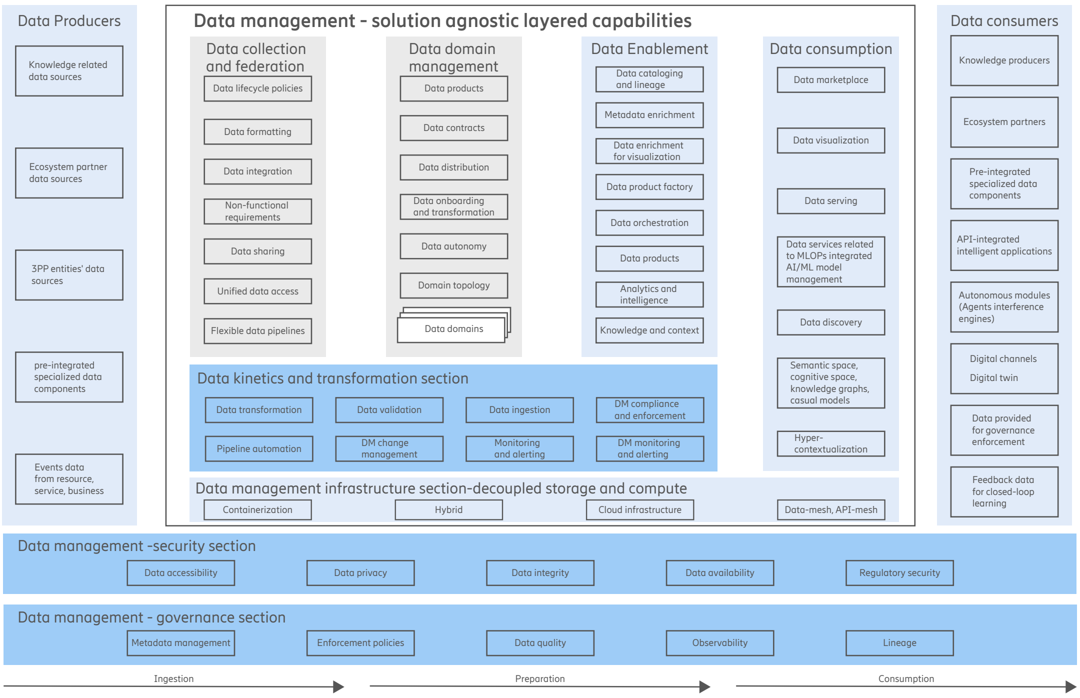
Data management systems are responsible for handling data from various sources. Once ingested, these systems become responsible for cataloging the data, applying de-identification and democratization principles, and refining raw datasets into processed and reliable ones. This makes the data ready for consumption by various consumers across a wide range of use cases.

Modern data management systems provide seamless integration with common data lakes with on-ramp and off-ramp functionalities that utilize data lakes as a long-term storage and analytical platform. Additionally, these systems are moving away from monolithic architecture to distributed federated systems, where compute resources can be positioned closer to the data. This approach leverages data federation to ensure data is always available for consumers.

These systems have to cover several use cases to securely manage data and to properly govern the capability to ensure data is processed and exposed efficiently and through common interfaces.

Some of the key features of modern data management systems are the following:

- data cataloging
- data quality reporting
- data lineage reporting
- data marketplace for data products
- data security and audit logging
- data ingestion frameworks
- data integration with data lakehouses
- data federation
- data classification, validation and transformations
- knowledge graphs & contextualization of data


Data Producers
Knowledge related data sources
Ecosystem partner data sources
3PP entities' data sources
pre-integrated specialized data components
Events data from resource, service, business
Data management - solution agnostic layered capabilities
Data collection and federation
Data lifecycle policies
Data formatting
Data integration
Non-functional requirements
Data sharing
Unified data access
Flexible data pipelines
Data domain management
Data products
Data contracts
Data distribution
Data onboarding and transformation
Data autonomy
Domain topology
Data domains
Data Enablement
Data cataloging and lineage
Metadata enrichment
Data enrichment for visualization
Data product factory
Data orchestration
Data products
Analytics and intelligence
Knowledge and context
Data consumption
Data marketplace
Data visualization
Data serving
Data services related to MLOPs integrated AI/ML model management
Data discovery
Semantic space, cognitive space, knowledge graphs, casual models
Hyper-contextualization
Data kinetics and transformation section
Data transformation
Data validation
Data ingestion
DM compliance and enforcement
Pipeline automation
DM change management
Monitoring and alerting
DM monitoring and alerting
Data management infrastructure section-decoupled storage and compute
Containerization
Hybrid
Cloud infrastructure
Data-mesh, API-mesh
Data consumers
Knowledge producers
Ecosystem partners
Pre-integrated specialized data components
API-integrated intelligent applications
Autonomous modules (Agents interference engines)
Digital channels
Digital twin
Data provided for governance enforcement
Feedback data for closed-loop learning
Data management -security section
Data accessibility
Data privacy
Data integrity
Data availability
Regulatory security
Data management - governance section
Metadata management
Enforcement policies
Data quality
Observability
Lineage
Ingestion
Preparation
Consumption

Data management systems follow a few core principles:

- **Data management architecture follows a federated approach:** the data ingestion architecture can easily scale up and down, support both batch and streaming data, and be flexible for different data types, sources or changing consumer scenarios.
- **Data is collected once, and many consumers are allowed:** federated systems have a data ingestion architecture that collects datasets only once, which is then made available for consumers with authorized access.
- Insights are shared: applications that produce insights can publish insights so other applications can also benefit from them.
- **Data is used in transparent, compliant, and ethical ways, with value for the end-user in mind:** data is democratized, meaning it is available to relevant consumers without compromising applicable security policies and regulations, as agreed by and between the data handler, the customer, employees, and partners.

Data islands can be considered so-called landing zones for data, where the data arrives, is processed, and exposed for near-real-time use cases. Data also needs to be stored for long-term use cases such as trend analysis and historical reporting. To achieve this, data islands must integrate with data lakehouses.

Target business objectives

This chapter covers the challenges associated with modernizing data architectures to meet the demands of autonomous network Level 5 state of autonomy in IT, network, and operations.

As networks evolve toward full Level 5 autonomy, where agentic AI systems perceive, reason, and act end-to-end with minimal or probably without human prompts, the data management future-state architecture must transform from batch-oriented stores into real-time, context-rich, AI-native fabrics. Below is a concise, structured view of the key challenges with integrating generative AI and agentic AI.

Challenges

Data fragmentation and federation:

- heterogeneous domain silos and multi-vendor requirements, such as RAN, core, edge, infra operations support systems/business support systems (OSS/BSS) data domains, impede seamless data unification
- complex federation protocols and privacy boundaries slow real-time data ingestion
- the system integrator (SI) drives brittle data management instead of industrialized and flexible data management

Real-time, low-latency processing:

- scaling streaming pipelines to handle millions of telemetry events per second
- avoiding so-called token overload in large language model (LLM) calls by dynamically curating context at inference time

Context management and relevance:

- balancing breadth versus depth by feeding agents only the precise briefing packet they need.
- preventing hallucinations and stale data drift in autonomous reasoning loops.

Governance, compliance, and security:

- embedding automated policy checks, personally identifiable information redaction, and ethical guardrails into every pipeline stage.
- ensuring auditable decision trails for every agentic action, across multi-party data sharing.

Observability and Agentic Operations complexity:

- instrumenting telemetry, logs, and feedback signals to monitor hundreds of collaborating agents.
- detecting and recovering from agent failures or adversarial inputs in real time

Model integration and life-cycle management:

- coordinating LLMs, small language models (SLMs), and specialized neural models across diverse data formats and service level agreements (SLAs).
- automating retraining, drift detection, and goal setting for evolving network conditions

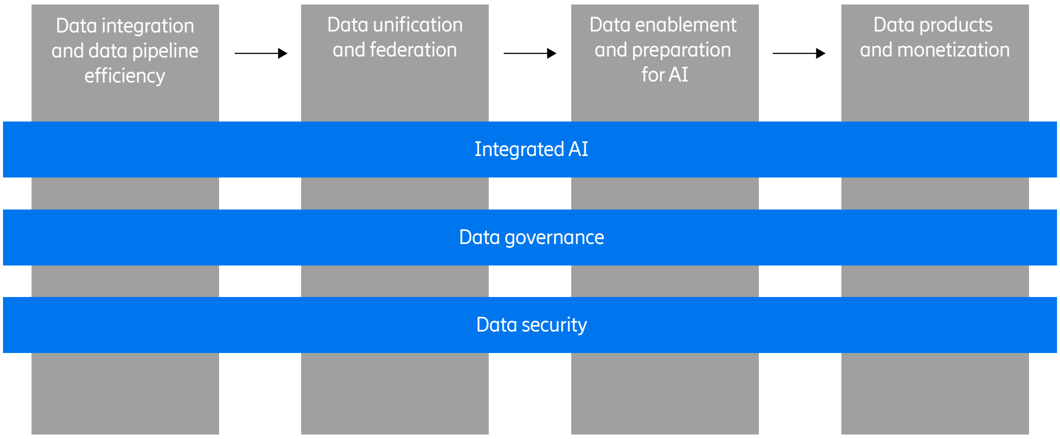
Data mesh and productization:

- enabling domain-oriented data-as-a-product with discoverable application protocol interfaces (APIs), to promote decentralization and ownership.
- leveraging semantic catalogs and knowledge graphs to enable contextual search for agentic AI agents.

North Star vision

The North Star vision for data management is the development of an evolved and AI-ready data management suite comprised of secure components, built on fundamental data management principles. It also supports data unification, federated data services, and extensible seamless integration with data sources.

All data is consumed in a homogenous manner regardless of the source, leveraging compound, well-governed, and efficient data pipelines serving AI-ready data with support for semantic modelling. This vision places particular emphasis on the different elements shown in the following diagram.


Data integration and data pipeline efficiency
Data unification and federation
Data enablement and preparation for AI
Data products and monetization
Integrated AI
Data governance
Data security

The grey columns represent the natural flow of data. Overlaying them are the data management aspects, marked in blue, which show the elements that are cross-cutting and intrinsic to each of the data management columns. These components should be regarded as fundamental to achieving the North Star vision.

Data unification and federation

Federated systems address many problems when it comes to proper scaling to increasing data volumes. Deploying compute resources closer to the data can optimize processing times and provide an efficient way to better use compute resources at smaller sites. For federated systems, it is critical to ensure a comprehensive and efficient data distribution framework that makes data accessible whenever needed.

These systems also require a common framework to support the creation and interworking of robust and seamlessly integrated data islands. These islands may consist of common data processing assets and differentiated assets. In addition, they implement a common specification for data management and sharing. Combining data islands into a mesh is a modern pattern for remote and distributed data management systems.

Data federation can be broken down into five main aspects:

1. A unified data catalog experience that allows consumers to discover all available data across all systems participating in the data mesh.
2. A cross-island pipeline orchestration, allowing consumers to request the collection or the production and exposure of data from remote systems.
3. A secure remote data access, where consumers are validated for trust, and the system is configured to expose only the necessary data elements.
4. Smart data movement to optimize latency and network traffic when multiple consumers have to gain access to remote data.
5. Data governance principles need to be realized on both local island and global mesh levels.

Data enablement and preparation for AI

In agentic AI and digital twin ecosystems, where autonomous, intelligent agents reason, act, and continuously learn from inputs, data enablement and being AI-ready become not only a precursor to analytics, but also emerge as a strategic capability. AI applications require reliable and timely access to trustworthy data that supports near-real-time decision-making, learning loops, and safe autonomy at scale, all of which are delivered through efficient, scalable AI-ready data pipelines.

As organizations accelerate the adoption of AI-ready data, the need to enable and manage such inputs semantically and contextually becomes paramount. In agentic AI, where data is consumed and actuated autonomously, feature engineering and feature stores must evolve beyond tactical AI or machine learning (ML) tooling into semantically enriched components, tightly integrated with semantic models, ontologies, and knowledge graphs. This will enable AI agents to interact, reason, explain, and act on data.

When feature engineering is introduced into the data pipeline, it consumes telemetry signals, pre-processes them, and creates features that are then placed in a feature store, becoming accessible for AI model training and inference. Semantic consistency is vital to ensure the agents interpret data contextually, preventing them from making incorrect assumptions, reasoning erroneously, or performing wrong actions. Semantic-aware feature engineering, which enriches the data pipeline with domain knowledge graphs, creates context-aware features by leveraging ontology-driven schemas and semantic layers for all data domains, making them contextually meaningful within a domain ontology or knowledge graph.

Adopting AI-driven automation for data enablement and preparation for AI allows adaptive data preparation loops. As AI models evolve and agents learn, data preparation must evolve too, addressing the auto-discovery of new, relevant features and integration with machine learning operations (MLOps) platforms to trigger re-training for continuous improvement. This ensures the co-evolution of data and agentic AI intelligence.

Data integration and data pipeline efficiency

Data integration

In today's telecom ecosystem, data flows from fragmented sources, whether it is legacy systems, edge devices, cloud-native services, internet of things (IoT) sensors, or partner APIs. Traditional, rigid integration models cannot accommodate AI-driven, real-time, and semantically enriched applications. Tomorrow's integration must be intelligent, adaptive, and autonomous, moving beyond data pipelines to orchestrate multi-source, multi-format, multi-speed flows across the full telecom stack.

North-south integration, reimagined: modern architectures treat both northbound systems, such as AI/ML, analytics, and digital apps, and southbound systems, such as networks, OSS/BSS, and edge, as intelligent, evolving entities in a continuously learning ecosystem. Semantic-aware connectors that auto-adapt to schemas, metadata shifts, and context changes will be essential.

AI-enhanced, composable pipelines: integration must become dynamic, composed by AI agents based on data needs and context. Using declarative metadata and policies, integration-as-code enables rapid onboarding of new domains, integration patterns, and external data products with minimal manual effort. For example, a context-aware AI agent could discover a 5G network event stream, infer its relevance for a predictive maintenance use case, and auto-integrate the stream with minimal human intervention.

Semantic interoperability: Future data integration frameworks embed semantic intelligence, leveraging ontologies and knowledge graphs to ensure that data is syntactically connected and contextually aligned. This enables cross-domain reasoning across RAN, core, OSS/BSS, and customer experience systems. Semantic mediation engines will map data from diverse domains into a unified knowledge model, enabling AI models to interpret and act with clarity, even in dynamic environments.

Autonomous integration services: AI-native integration frameworks are equipped with self-healing, self-optimizing capabilities, observing the telemetry of integration pipelines to perform the following:

- auto-retry or re-route failed data streams
- trigger schema reconciliation workflows
- optimize data flow paths for latency or cost
- auto-discover new sources through metadata harvesting and usage analytics

Edge-to-cloud integration: data integration should support both edge and cloud environments. This allows fast and efficient data sharing from edge devices, while also ensuring the data remains consistent for central analytics.

Using distributed microservices and streaming architectures, data can be processed and cleaned close to where it is generated and then send the data's useful parts to cloud systems. This setup also allows AI at the edge to work with local data, while still contributing to a larger, connected intelligence system.

Strategic North Star: the ultimate vision is a fully abstracted and policy-driven data integration fabric, where data consumers simply define what they need, and the system intelligently orchestrates how to acquire, transform, and deliver it, doing so contextually, securely, and in real-time. This aligns seamlessly with a data-mesh philosophy, in which integration is decentralized but governed, semantically enriched, and optimized for agility, scalability, and trust.

Data pipeline efficiency

Data pipelines are built to transit and process large, disparate datasets, and can commonly be understood as following extract, transform, load (ETL) processes. However, that is a simplification; more often, a pipeline is defined as an automated process for the continuous collection and exposure of data.

The first step toward building efficient data pipelines starts with observability. It is important to have telemetry data at every step of the pipeline. Industry standards revolve around the collection of metrics, logs, and traces with many observability frameworks, such as OpenTelemetry, allowing a common framework for telemetry collection and an emerging standard for telemetry syntax. The telecom industry has typically added a fourth telemetry signal in fault indicators, or more commonly referred to as alarms. Alarms are essentially pre-created insights that indicate a potential issue in the running of an application. But in cloud-native observability, post-hoc analysis of logs to detect faults is highly recommended, as it simplifies the application software responsibility from self-analysis to just producing raw telemetry.

Every data collector, data processor, message bus, object store, and database in the data management system must report telemetry, such as processing rate, processing volume, logging, and metrics for failures, in the processing system.

A centralized observability function observes the telemetry coming from all the components of the data pipeline and can take control to reconfigure it. The function may decide to create priority flows for certain types of data in the pipeline or modify data retention to avoid depleting available storage. The function could additionally observe increasing latency and decide if it needs to orchestrate more instances of a collector or message bus broker to increase the throughput.

This use case of automated pipeline management and efficiency monitoring lines up with an agentic architecture. AI can create insights from the telemetry of your data pipeline, where an agent processes insights and maps to actions, while another agent takes responsibility for applying actions and monitoring the effect of the action on the system.

Data governance

As telecom operators strive to monetize data, improve customer experiences, and comply with increasingly strict regulatory frameworks, effective data governance has become a foundational capability. It requires an integrated approach to governance that includes data ingestion, storage, processing, sharing, and analytics. This means that capabilities need to be considered across the data management architecture to ensure that policies, security, and privacy are continuously followed.

Data security

Privacy-aware data de-identification and unification

Addressing privacy and data sensitivity with proper privacy-preserving techniques for anonymization, pseudo-anonymization, and unifying data from silos may result in re-identification risks, violating data sovereignty rules, or breaching compliance.

In the shift toward agentic AI ecosystems, data security is no longer about perimeter defense; it becomes more semantic, dynamic, embedded, ontology and agent-aware.

Data security needs to travel with the data, be interpretable by agents, and adapt in real-time to context. This requires deep integration with data fabrics, semantic models, policy engines, and identity frameworks to ensure that autonomy and security can grow in tandem.

Classification and AI-enabled access control and de-identification

An important part of securing data is understanding its nature and adhering to the principles that define how to handle different data types. AI agents can assist in classifying data based on agents trained on security standards such as the European Union's General Data Protection Regulation (GDPR). As part of the classification process, the data is tagged and configured to correspond to the data handling policies for the rest of the data pipeline.

A current industry trend is compliance as code, a practice of specifying policies as code, which allows adding policy decisions into a runtime policy engine for continuous access control verification.

In federated systems, it is important to ensure that access control persists when data is moved. Therefore, as part of the data transfer, the data object that gets transferred contains not just the information elements of the data but also the metadata that instructs the receiving system on the correct access control configuration for the data once transferred. This includes role-based, attribute-based, and policy-based access control information elements to provide data security in the receiving system.

Data products and value

Address the creation and exposure of data products and the opportunities for communication service provider (CSP) monetization not only on a product level but also at the capacity level within a data product.

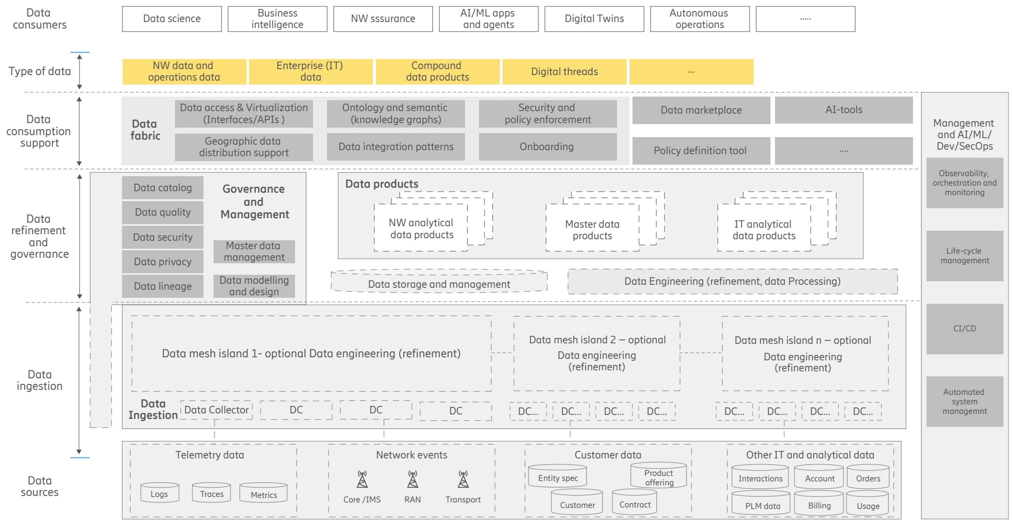
Data products are valuable and actionable data sets offered to data consumers by the data management system. These products can be aggregated and processed into datasets or insights generated from the collected data. Data products are traceable and trustworthy, meaning there is an associated data lineage with the product, which details every processing step that has been executed in the creation of that product, and what datasets it has been derived from. In addition, the data product should have a reported data quality, which is a calculated indicator comprised of data completeness, accuracy, consistency, and validity.

For monetization purposes, data products can be offered through a data marketplace, advertising their availability. This type of marketplace also offers the ability to configure the exposure of data to consumers who purchase access to the data product.

Additionally, monetization can be offered on solutions that leverage data products and insights for autonomous networks, network planning, and trend reporting.

Reference architecture

This section illustrates the future-state reference architecture of AI-ready data management. This architecture mainly consists of four areas: data sources, data ingestion, data refinement and governance, and data consumption support.


Data consumers
Data science
Business intelligence
NW sssurance
AI/ML apps and agents
Digital Twins
Autonomous operations
.....
Type of data
NW data and operations data
Enterprise (IT) data
Compound data products
Digital threads
...
Data consumption support
Data fabric
Data access & Virtualization (Interfaces/APIs)
Geographic data distribution support
Ontology and semantic (knowledge graphs)
Data integration patterns
Security and policy enforcement
Onboarding
Data marketplace
AI-tools
Policy definition tool
....
Management and AI/ML/ Dev/SecOps
Observability, orchestration and monitoring
Life-cycle management
CI/CD
Automated system managemnt
Data refinement and governance
Data catalog
Data quality
Data security
Data privacy
Data lineage
Governance and Management
Master data management
Data modelling and design
Data products
NW analytical data products
Master data products
IT analytical data products
Data storage and management
Data Engineering (refinement, data Processing)
Data ingestion
Data mesh island 1- optional Data engineering (refinement)
Data mesh island 2 – optional
Data engineering (refinement)
Data mesh island n – optional
Data engineering (refinement)
Data Ingestion
Data Collector
DC
DC
DC
DC...
DC...
DC...
DC...
DC...
DC...
DC...
DC...
Data sources
Telemetry data
Logs
Traces
Metrics
Network events
Core /IMS
RAN
Transport
Customer data
Entity spec
Product offering
Customer
Contract
Other IT and analytical data
Interactions
Account
Orders
PLM data
Billing
Usage

Standardization activities

Ericsson's goal is to avoid fragmentation and promote industry alignment. The reference architecture described herein is based on standards from the 3rd Generation Partnership Project (3GPP), Open RAN (O-RAN), and open data lakehouse Apache projects. Extensions and new features introduced in the reference architecture will be fed into the impacted standardization bodies with a clear goal to harmonize architectures and solutions.

Conclusion

A data foundation is a key facilitator for AI, analytics, and automation. It must be multi-domain and adaptable to meet different consumer needs. Establishing a data foundation is essential to succeed in the digital transformation, where AI will play an important role. Many AI initiatives fail because this foundation is not strong enough.

This document describes a reference architecture that provides a future-proof data foundation that scales elastically, supports seamless data integration, can be deployed in a hybrid manner, and is purpose-built for AI-native intelligence, autonomous network operations, and industry alignment.

Authors

Bo Åström is an expert in system and service architectures. He joined Ericsson in 1985 and has extensive experience in radio networks, core networks and enterprise service networks, where he has worked with standardization and product development. Earlier in his career, Åström held technology specialist roles in the areas of interfaces and protocols, messaging architectures and network architectures. He holds more than 70 patents.

Bulent Gecer joined Ericsson in 2001. He is an expert in data management and incident analytics within Ericsson's CTO office and has played a leading role in developing data-management- and analytics-related architectures at corporate level for the past 15 years. Gecer previously worked as a technology specialist and software developer within core networks. He holds an M.Sc. in engineering physics from Uppsala University in Sweden.

Anna-Karin Rönnberg is an AI and data evangelist with broad experience ranging from systems management to portfolio management. She joined Ericsson in 1985 and currently serves as a portfolio manager within the CTO office and as such responsible for strategy and portfolio execution. In recent years, Rönnberg's work has focused on enabling a more data- and AI-driven strategy and portfolio.

Michael Buisman is an expert in data management, CEM and AI with over 25 years experience in Telecoms. Michael joined Ericsson in 2007 and has experience in both Managed Service Networks and Core Networks in developing and deploying Automation, Analytical, Data and AI solutions.

Soren Marklund joined Ericsson in 1986. He is advising Ericsson's global services' data-driven service transformation. He has extensive experience in innovating and deploying customer-centric data-driven solutions, having held sales, services, operational, data, and strategy positions globally. Frequently engaged in customer and industry sessions around Data and Agentic AI strategies.

James Dwyer joined Ericsson in 2017. He is the Data Management Architect for the network management portfolio. In Ericsson, he was worked extensively in designing event driven systems, specializing in the collection, processing & exposure of observability data. Recently, he has focused on contributing to the standardization of the O-RAN Data Management & Exposure function of the SMO along with it's implementation in Ericsson.

Neeraj Joshi is a senior product management leader working at the intersection of technology, strategy, and data driven transformation. He helps CSPs unlock value from data, accelerate automation, and prepare for AI-powered autonomous networks of the future. Specializing in telco data management, DataOps platforms, and cross-domain data integration and transformation, he has spent nearly two decades at Ericsson, turning complex visions into meaningful, scalable business outcomes. Earlier in his career, he led solution design and product strategy across service enablement platforms, machine-to-machine/IoT systems, and industry-vertical solutions spanning healthcare, utilities, transportation, and mobile advertising.

Richie Dalton is a product management leader working in Solution Area Network Management. He joined Ericsson in 1998 and worked in Network Management and OSS Engineering roles until 2024, specialising as an Expert in Network Automation and Evolved Network Management. Since transitioning to product management, he has taken responsibility for portfolio strategy and Data Management across the SA NM portfolio.

Adam Bergkvist joined Research Ericsson in 2007. He has a background in service prototyping, open source development and web standardization. In recent years, his areas of research have been cloud computing and data infrastructure. He holds an M.Sc. in software systems engineering from Luleå University of Technology in Sweden.