The role of knowledge and data services in unified networking experience technology (UNEXT)

White paper

Alberto Conte, Senior Research Specialist

To give a simple and secure experience to anyone that uses or interacts with the network, from end user to developers, enterprises and network operators, we need to look at the networking in a fundamentally different way. We call this approach unified networking experience (UNEXT). Through autonomous management, secure and trusted interactions across stakeholders, and seamless orchestration, the UNEXT unified system integrates innovative capabilities from five diverse research areas: autonomous services, decentralized environments, extended computing services, network-application symbiosis, and knowledge and data services (KDS), the latter being the subject of this white paper.

The overall ambition of KDS is to provide the “nervous system” of UNEXT by underpinning the collection, exposure, discovery, sharing, composition, and exploitation of knowledge. The KDS goal is to enable multiple actors, who may not all be trustworthy, to share, discover, access, and utilize knowledge coming from heterogenous sources. This whitepaper describes the context, challenges and proposed approaches related to the KDS research area.

Contents

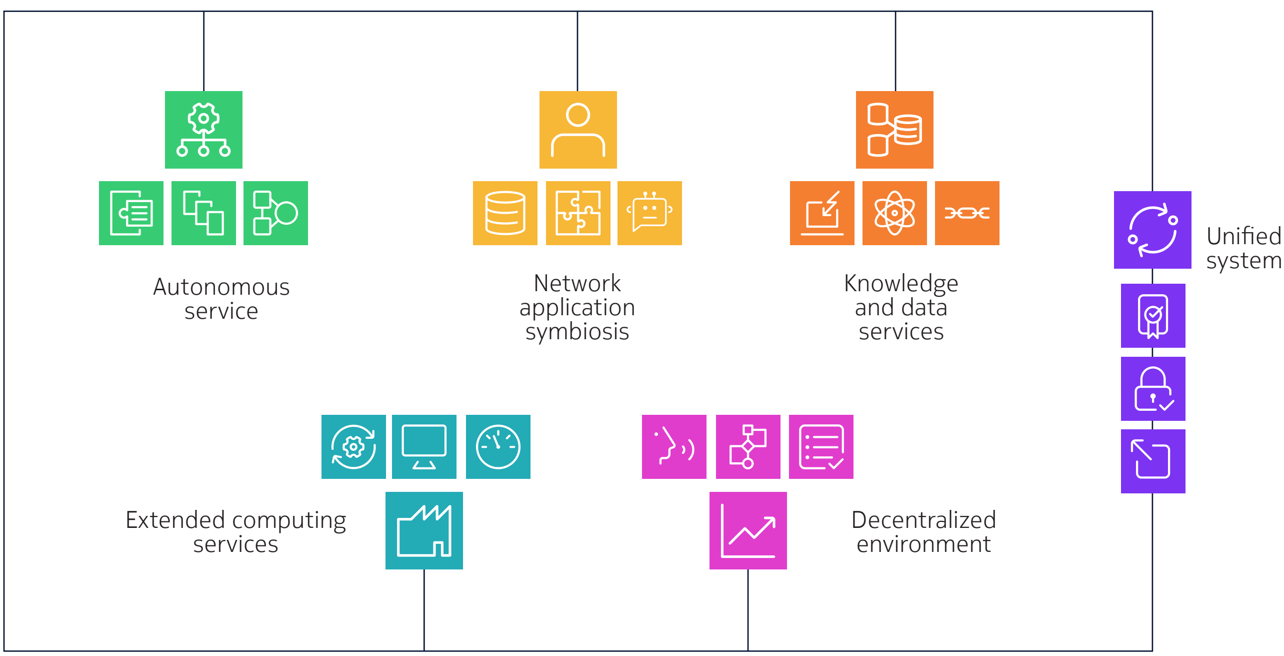
NOKIA BELL LABS

Introduction

In the complex future of a multi-player, digital-physical world with increasing security threats, UNEXT [1] proposes a novel approach to realize simple, secure and scalable networking services based on composable software elements. The intrinsic capabilities of UNEXT are defined by modules that implement its constituent elements and functions, such as network and compute resources, data collection, storage and sharing modules, automation closed loops and cognitive agents. The capabilities exposed to different users and ecosystem partners in the value chain can be custom-combined to offer more versatile, flexible and individually optimized solutions.

UNEXT will achieve this through autonomous management, secure and trusted interactions between stakeholders, and seamless orchestration. To accomplish this ambitious goal, the UNEXT unified system integrates innovative capabilities from five different research axes: autonomous decision-making and service realization, decentralized coordination among heterogeneous stakeholders, extended computing at the edge, network-application co-optimization, and the focus of this whitepaper, knowledge and data services (KDS).

Figure 1. UNEXT and its constituent technical domains



This whitepaper describes the context, challenges and proposed approaches related to the KDS research area. KDS deals with the evolution of communication systems towards intent-based, data-driven and AI-native systems. Such systems rely heavily on the availability and accessibility of relevant data and knowledge, provided in a timely manner and at the right place. In this context, KDS defines the key principles for an evolutive knowledge fabric, integrated at the core of UNEXT and able to accommodate novel cognitive techniques that may emerge in the future. Metaphorically, the ambition of KDS is to provide the nervous system of UNEXT, underpinning the collection, exposure, discovery, sharing, composition, and exploitation of knowledge. Its goal is to enable multiple actors, who may not all be trustworthy, to share, discover, access, and utilize knowledge with one another.

Knowledge and data services

Why a knowledge system is necessary

A foundational requirement for data-driven and AI-native systems is the availability and accessibility of relevant data and knowledge, provided in a timely manner and in the right place, while adhering to access, privacy and sustainability constraints. Such knowledge is made available to and by cognitive agents performing advanced reasoning on the system.

Here, the term knowledge is used as an umbrella term for the vast and diverse array of information involved, including data, whether raw (from sensors, monitoring, measurements, logs, etc.) or processed (statistics, analytics, predictions, etc.), events (alerts, threshold triggering, etc.), knowledge corpuses (ontologies, specifications, etc.) lessons learned (e.g., history of past state-action-reward), ML models, and “recipes” to problem-solving (blueprints, pipelines). In a few words, knowledge is everything needed to understand, reason about and act upon what is happening in the system.

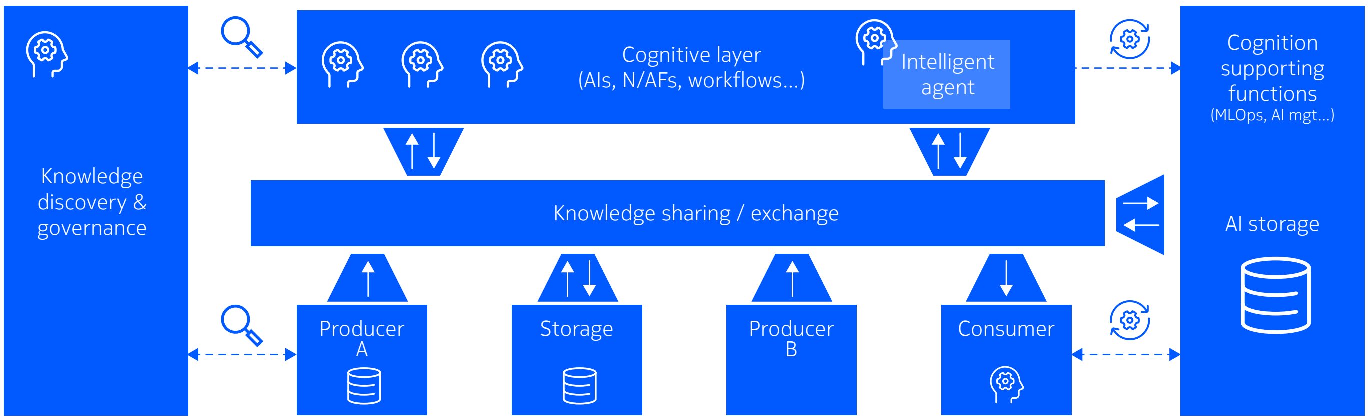
In the early 2000s, Clark, D.D. et al. developed the concept of the ‘knowledge plane’ [2] capturing and formalizing the necessity of gathering and making knowledge available in the networks. With the advent of machine learning and AI, the concept evolved further into a knowledge system integrating functions that support cognition. The goal of a knowledge system is to enable multiple producers, potentially from different administrative domains, to make their knowledge available to cognitive consumers. Consumers can, in turn, become producers by re-injecting their processing results into the system.

Different approaches can realize knowledge systems to satisfy specific requirements. For example, designing the system for a single domain, isolated from the external world, is very different from targeting a multi-stakeholder environment. However, it is possible to sketch the generic structure of any knowledge system as composed of five main logical functionalities:

1. Knowledge production and storage
2. Knowledge discovery
3. Knowledge sharing/exchange
4. Cognitive processing
5. Knowledge governance

Figure 2 provides a logical (and simplified) schematic of a knowledge system. Two producers (A and B) make available their knowledge: one (B) by storing it in a storage system (e.g., a data lake), the other (A) by relying on its own internal storage. Consumers can use discovery and recommendation mechanisms to find the producers that best satisfy their needs under the control of governance rules. Producers and consumers securely share the selected knowledge through adapted sharing and exchange mechanisms (pub-sub, querying, etc.). It is then processed by cognitive consumers, such as network/application functions (N/AF), intelligent agents or more complex data pipelines and cognitive workflows, assisted by MLOps and AIOps supporting functions (e.g., training facilities).

Figure 2. Simplified schematic of a knowledge system



Challenges, requirements and trends

Realizing a knowledge system is not just a matter of selecting and integrating the most suitable tools and solutions available from the plethora of offerings in the ecosystem. Sharing knowledge faces complex challenges, particularly in the context of decentralized multi-stakeholder environments, which are the natural target for UNEXT [3]. These challenges include the following:

- Difficulty in accessing and sharing data and knowledge across administrative boundaries due to privacy, trust and monetization concerns
- Incomplete, inconsistent, biased, conflicting, misleading and even malicious information from partially trusted or untrusted actors
- Interconnection and interoperability issues due to different sharing solutions, with different operating schemes and heterogeneous data and knowledge formatting
- Integration of (far) edge elements such as mobile phones and IoT devices into the knowledge production process along with specific execution environments such as AI accelerators
- Cooperation and composition among cognitive processes and agents distributed across multiple domains
- Unsustainable growth of data and knowledge exchanges also known as “data storms” that negatively impact system performance, costs and carbon footprint.

Simultaneously addressing all the requirements and challenges remains an open issue. Mainstream solutions opt for centralization, while a few others propose alternative paths. While the goal of this document is not to provide a survey of all the possible approaches, we will analyze some notions and trends relevant to building the UNEXT KDS vision.

Concerning storage, the trend in the last decade has been centralization, which evolved from data-warehouses storing pre-structured data based on the extract-transform-load (ETL) paradigm, to unstructured data lakes following the extract-load-transform (ELT) paradigm and data-fabrics aiming at unified management [4]. Centralized approaches are effective in gathering and exposing a global knowledge base to multiple cognitive agents but struggle with decentralization, locality, privacy, and trust aspects. New approaches, such as in distributed data marketplaces, data mesh [4], [5], or the extreme database-per-service pattern [6], challenge the fully centralized path by allowing producers to store and manage data locally. These evolutions aim at better supporting data decentralization; each producer oversees the management, exposure, and governance of its own data. Compatibility and scaling are the major challenges of these approaches.

Knowledge discovery can be realized using two approaches. In the expose-discover approach, producers publish what they can provide into logically centralized knowledge catalogs, from which consumers discover and select the best producers for their needs. In the request-offer approach, consumers express their needs ("this is the information I need"), and the producers can satisfy them (totally or partially) by offering their services, with associated information to help in selection (cost, accuracy, etc.).

Knowledge sharing and exchange connect all the corners and layers of a network and provide seamless access to the required knowledge [7]. Although the abstraction level, structure, and format of the knowledge resources vary, all are represented as chunks of bits, which are then made available and transferred to consumers in a push or pull mode. In pull mode, the content is fetched by the receiver every time it is needed. In push mode, a distribution agent is instructed to forward the bit chunks to interested consumers.

Another dimension of the communication pattern is the number of participants, scaling from one-to-one (producer-consumer direct exchange) to one-to-many (called subscribe-notify in the streaming case) and many-to-many as in publish-subscribe (pub-sub). Pub-sub decouples communicating endpoints in space, time, and synchronization and enables simultaneous asynchronous sharing of the content with multiple consumers. The producer publishes the data only once, and the pub-sub makes it available to all interested consumers. Irrespective of the sharing patterns, the current approach of collecting and centralizing all available data and knowledge is problematic in terms of sustainability and costs, which argues for rationalizing production and collection (e.g., through programmability). Finally, when it comes to the integration of AI/ML into the knowledge system, one of the key elements is the support of their specific operations (AI/MLOps): model management, training, validation, conversion, inference, monitoring, explainability, etc. While many solutions are available in the ecosystem, moving away from current per-use-case ad-hoc operations to full automation is still a challenge.

Central ML Platforms (MLPs), such as the ones proposed by webscalers, offer powerful tools but are not well adapted to distributed and decentralized environments. The flexibility to move from one MLP to another appears to be an important requirement to avoid vendor lock-in and to enable the use of local-specific platforms and specialized hardware, especially when considering far-edge environments that may offer specific architectures and/or limited compute resources. Another key aspect to consider when designing the cognitive system is the continuous emergence and evolution of AI techniques such as large language models (LLMs), introducing new operations such as retrieval-augmented generation (RAG) and prompts, and operational models (whitebox vs. blackbox, LLM agents, etc.). With all the AI-related operations in place, a cognitive system facilitates the cooperation and composition of multiple cognitive agents and the orchestration of complex data pipelines and reasoning processes.

Examples and use cases

The need for knowledge systems exists in today's networks and is expected to emerge in new forms as communications ecosystems evolve. Some examples of these are provided here, illustrating environments to which the developed technologies could be applied.

- **Analytics in mobile networks:** The network data analytics function (NWDAF) [8]) is a 5G core network function defined by 3GPP that enables network data analysis. It allows consumers to obtain data, statistics, and predictions on network and service metrics (e.g., the evolution of the load of a network function). NWDAF centralizes data collection, ML training, and inference in the 5G core network and can interact with other NWDAF instances. Further evolutions in 5G Advanced and 6G will integrate new features such as federated learning and radio access network (RAN) analytics.
- **End-to-end slicing/service assurance:** When providing end-to-end services such as network slicing guaranteeing a service level agreement (SLA) is crucial. It requires the optimization and calibration of network performance and operations amid growing complexity and interconnectivity. The challenges include diverse data sources, use-case-specific KPIs, domain-specific measurement metrics, and vendor-specific formats and protocols.
- **Cognition at the edge:** Leveraging increasingly abundant and capable compute resources at the edges of networks minimizes latency and cost and maximizes performance and sustainability [9]. This opens up the opportunity to execute cognition processes closer to applications and data sources, while meeting the challenges associated with privacy and specific software and hardware execution environments.
- **Public cloud AI/ML services:** Commercial cloud providers offer ML platforms (MLPs) and frameworks as a service, supporting the development and execution of AI/ML-based solutions. Specifics of the platforms and frameworks can differ among public cloud providers, thus requiring AI/ML-based solutions running in them to be compatible or adaptable to the diverse environments.

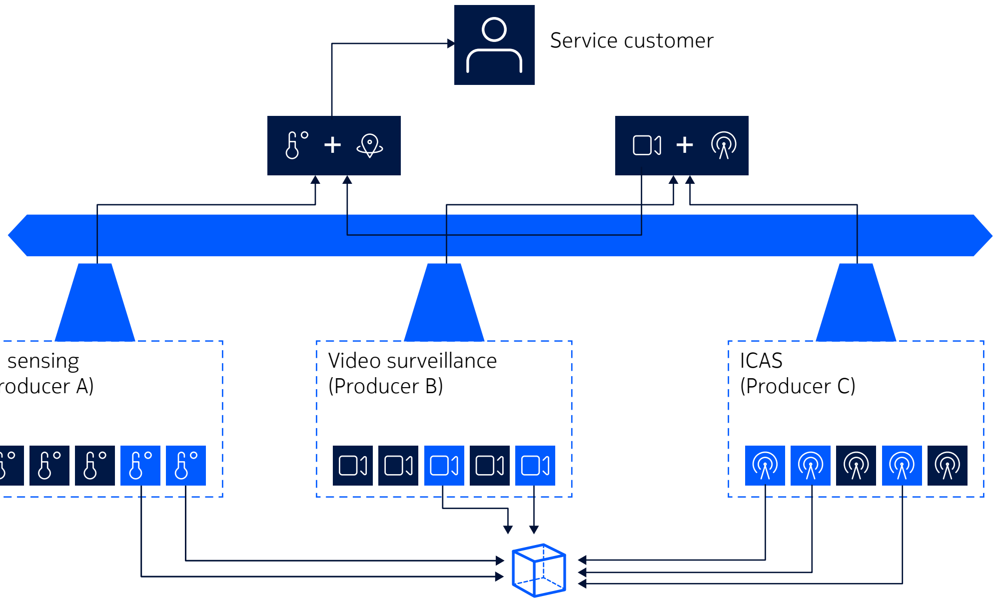
These examples illustrate how the complexity and variety of knowledge-related operations already necessitate flexible and evolutive solutions in today's networks. The following use cases complement this view with a look to the future, describing two hypothetical services and the challenges faced in realizing them in highly automated and decentralized environments.

Enhanced multi-modal sensing system

In this use case (figure 3), multiple independent sensing systems (producers A, B and C), managed by different stakeholders, cooperate to provide rich information on the monitored environment. As an example, we can imagine combining the integrated communication and sensing (ICAS) [10]) radar-like capabilities with video surveillance and temperature sensing. Such a multi-modal system would be able to detect and locate objects as well as characterize them in terms of temperature, shape and type. As an operational use case, we can imagine such a system in use in a factory to detect and locate objects of a specific size (e.g., >5 cm) that may inadvertently drop on the autonomous guided vehicle (AGV) or robot corridors of a production floor. The combined information may be used by the factory automation system to rapidly react to the event, for example, by rerouting the AGVs or robots while selecting the best recovery action: send a human operator or, alternatively, use a specialized robot in case the object is too heavy, big or hot.

The challenges here come from the difficulty of sharing knowledge across multiple actors as well as the necessity to minimize the collected data for performance and sustainability, for example, by dynamically calibrating the sensing granularity.

Figure 3. Multi-modal sensing system



Hidden correlations

This use case investigates the challenge of detecting hidden or unexpected correlations. Because of a malfunctioning mechanism, a window in a server room remained open during a heavy rain at night. Some water spilled on a server, provoking a major failure. While detecting the cause of the server failure (an open window on a rainy night) would be trivial for a human, it proves to be extremely difficult in highly automated, unmanned environments, composed of multiple independent systems, such as smart building piloting windows, meteorological sensing and IT management.

In this use case, the key challenge is to detect the (hidden) correlation between events that weren't expected to be correlated. A related risk may arise from the adoption of a naive collect-and-analyze-all approach that entails exponential processing costs.

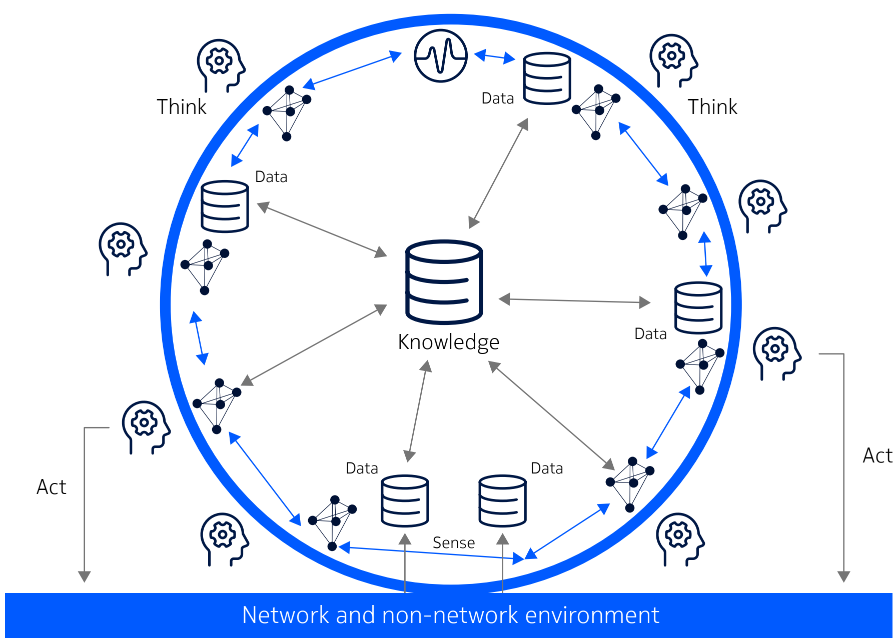
Exploring semantic proximities and exploiting findings from other systems may prove to be more efficient and sustainable.

KDS approach and directions

KDS specializes the abstract concept of knowledge system presented above with UNEXT-specific working assumptions: decentralization, autonomous multi-agents, intent-based operations, compute continuum, security, and sustainability. The ambition is to provide the nervous system of UNEXT, supporting knowledge collection, exposure, discovery, sharing, composition and exploitation.

In a nutshell, the KDS approach is to move away from centralized data storage and analytics to a system where any domain or agent can offer its data/analytics services and benefit from the knowledge capabilities of other domains and agents, irrespective of their location, execution environment and ownership. The goal is to enrich local cognition through dynamic and opportunistic (loose) cooperation among intelligent autonomous agents based on their shared respective knowledge (Figure 4).

Figure 4. KDS enables sense-think-act cooperation between intelligent agents



Realization of the KDS vision is based on four foundational pillars:

1. Knowledge continuum
2. Intent-based semantic discovery
3. Programmable knowledge production and exchange
4. Cross-domain cooperative cognition

Knowledge continuum

Acknowledging the multi-stakeholder and massively distributed/decentralized nature of UNEXT, KDS proposes the concept of knowledge continuum. Extending the concept of compute continuum, which aims at interconnecting multiple (network-) compute resources belonging to different administrative domains [11], the knowledge continuum aims at interconnecting multiple administrative knowledge domains and agents to enable the extraction and sharing of knowledge and the operation of complex cognitive workflows across highly heterogeneous infrastructures. Thus, the knowledge continuum spans across the whole end-to-end system, including extreme-edge elements (e.g., mobile phones, IoT devices, vehicles, etc.) as well as specific execution resources (e.g., hardware accelerators dedicated to AI and ML).

In the KDS vision, each UNEXT domain or agent has full autonomy and responsibility for realizing its internal knowledge system with associated governance, but it must comply with KDS discovery and access principles when participating in the knowledge continuum. The latter is achieved through a set of cross-domain functions supporting the interconnection of the knowledge systems and ensuring semantic continuity and coherency across them. Similarly, AIOps support functions and resources are federated and made available across domains. With the ability to discover and access knowledge everywhere, it becomes possible to build and orchestrate complex cognitive pipelines across the whole continuum.

Intent-based semantic discovery

Starting from the simple consideration that knowledge is worth nothing unless we know what it means, KDS proposes to rely on semantic interfaces to describe the available knowledge and to leverage recent progress on natural language processing (NLP) and generative AI (GenAI) to enable intent-based discovery. In this approach, consumers express their needs in a declarative way (e.g., "I need meteorological information about New York City"), and the discovery mechanism uses semantic proximity to select the best producers (e.g., producer A provides the temperature, producer B provides the humidity measures).

The goal is to introduce further de-correlation between consumers and producers. With intent-based semantic discovery, the consumer no longer needs to know what knowledge is available in the system; the nervous system oversees the selection (and update) of the best possible match, depending on the current knowledge availability. Producers can now easily join or leave the system. Semantic-based operations ease the emergence of knowledge recommendation systems and are a cornerstone in building semantic continuity solutions that preserve, recover, and even discover (hidden) correlations in knowledge coming from distributed domains.

Semantic discovery is based on the request-offer paradigm described in the "Challenges, requirements and trends" section above.

Programmable and efficient knowledge production and exchange

A key goal of the UNEXT nervous system is to make relevant knowledge available in time and at the right place. This can be realized with programmable knowledge production and exchange services. On the one hand, the granularity of knowledge production shall be dynamically tuned to match the specific consumer needs and to avoid wasting resources. Network load, privacy, monitoring targets and energy efficiency are some examples of criteria to consider when programming and tuning the production. On the other hand, exchange mechanisms shall offer multiple speed lanes, dynamically selectable to satisfy the specific needs of each exchange (e.g., low latency, data bulk, multicast, etc.). The characterization of knowledge exchanges enables this, as it allows us to automatically "classify" each exchange in terms of needs and dynamically adjust to their changes.

Overall, the target of knowledge programmability is to move toward knowledge frugality for system performance/efficiency and sustainability reasons.

Cross-domain cooperative cognition

In the KDS vision, multiple autonomous cognitive agents interact to understand, reason about and act upon what is happening in the system. Dealing with the proliferation of AIs requires full automation of the cognitive system. The intention is to move away from the current per-use-case ad-hoc processes toward an AI “plug-and-play” approach. Two directions must be developed: the orchestration and operation of AIs across the knowledge continuum and the composition of modular AIs into complex cognitive workflows.

Concerning AI orchestration and management, KDS targets the portability of AI operations across the knowledge continuum, recognizing the need and efficiency of exploiting local and distributed AIOps resources (instead of building yet another centralized ML platform). This allows for easy integration of novel operations, for example, to support LLM (LLMOps) and distributed and federated learning (FLOps). In addition to supporting general operations, each domain can make available, when necessary, specific operations such as AI conversion to some specific execution environment. KDS also advocates for a tighter integration of AIOps, AI monitoring and intent-based AI orchestration.

The second direction is about the composition of complex cognitive workflows. The challenges reside in the ability to tailor existing solutions to new use cases by evaluating source-target use case “proximity” and then, based on that, by selecting and sequencing the right set of cognitive modules (full AIs or smaller components) together with the necessary adaptations (e.g., fine tuning, conversions, etc.).

Conclusions

The spectacular progress of AI technologies in recent years is driving the emergence of a new generation of AI-native networking systems, integrating and interconnecting advanced cognitive and analytics capabilities, and offering unprecedented automation and adaptability.

The technologies developed by KDS aim to provide the “nervous system” of such new networking systems, enabling the collection, exposure, discovery, sharing, composition, and exploitation of knowledge across a decentralized, multi-stakeholder environment.

We have highlighted the challenges to be tackled when building AI-native systems, including data access limitations, information quality concerns, interoperability issues, and the need for sustainable knowledge management. To address these challenges, KDS proposes a novel approach based on four foundational pillars:

1. The ability to operate cognitive processes all along the knowledge continuum
2. Intent-based interactions over semantic interfaces
3. Efficient knowledge production and exchange through programmability and dynamic tuning
4. Cross-domain cooperative cognition, with security and sustainability acting as transversal needs.

By embracing these principles, KDS aims to move away from centralized data storage and analytics towards a system where any stakeholder or agent can offer its data and analytics services and benefit from the collective knowledge of the networking system. This approach fosters dynamic and opportunistic cooperation among intelligent autonomous agents, ultimately enriching local cognition and enabling the realization of sophisticated, data-driven applications in the UNEXT ecosystem.

The KDS vision represents a significant step towards a future where communication systems are more intelligent, adaptable, modular and resilient, capable of leveraging the vast potential of knowledge to address the challenges of a complex and interconnected world.

Abbreviations

AGV	Autonomous guided vehicle
AI	Artificial intelligence
AIOps	AI operations
ELT	Extract-load-transform
ETL	Extract-transform-load
FLOps	Federated learning operations
GenAI	Generative AI
ICAS	Integrated communication and sensing
KDS	Knowledge and data services
KPI	Key performance indicator
LLM	Large language model
LLMOps	LLM operations
ML	Machine learning
MLOps	ML operations
MLP	ML platform
N/AF	Network/application functions
NLP	Natural language processing
NWDAF	Network data analytics function
Pub-sub	Publish-subscribe
RAG	Retrieval-augmented generation
RAN	Radio access network
SLA	Service level agreement
UNEXT	Unified networking experience technology

References

[1] Sefidcon, A., Vulkán, C., and Gruber, M., "UNEXT: A unified networking experience," Nokia Bell Labs whitepaper, 2023. Online: https://www.bell-labs.com/institute/white-papers/unext-a-unified-networking-experience/#gref

[2] Clark, D. D., Partridge, C., Ramming J. C. and Wroclawski, J., "A knowledge plane for the internet", SIGCOMM 2003, 2003. Online: https://groups.csail.mit.edu/ana/Publications/PubPDFs/A%20knowlege%20plane%20for%20the%20internet.pdf

[3] Atkinson G. et al., "A unified networking experience in decentralized environments", Nokia whitepaper, 2024. Online: https://onestore.nokia.com/asset/214086

[4] Bluhm, I. et al., "Data products, data mesh, and data fabric," Bus Inf Syst Eng, 2024. Online: https://doi.org/10.1007/s12599-024-00876-5

[5] Dehghani, Z., "Data Mesh: Delivering Data-Driven Value at Scale", O'Reilly Media, Inc., 2022.

[6] C. Richardson, "Pattern: Database per Service," Microservice.io web site, 2024. Online: https://microservices.io/patterns/data/database-per-service.html

[7] Ziegler, V. et al., "6G Architecture to Connect the Worlds," in IEEE Access 8, pp. 173508-173520, 2020. Online: https://ieeexplore.ieee.org/document/9200631

[8] 3GPP, "TS 23.501: System architecture for the 5G System (5GS)," Rel19, Jun 24. Online: https://www.3gpp.org/ftp/Specs/archive/23_series/23.501/23501-j00.zip

[9] "Edge Computing: The Fourth Wave Rises", Moor Insights & Strategy, May 2018. Online: https://moorinsightsstrategy.com/wp-content/uploads/2018/06/Edge-Computing-The-Fourth-Wave-Rises-By-Moor-Insights-And-Strategy.pdf

[10] Wild, T., Braun V. and Viswanathan, H., "Joint Design of Communication and Sensing for Beyond 5G and 6G Systems," IEEE Access 9, pp. 30845-30857, 2021. Online: https://ieeexplore.ieee.org/document/9354629

[11] Kokkonen, H. et al., "Autonomy and Intelligence in the Computing Continuum: Challenges, Enablers, and Future Directions for Orchestration," 2023. Online: https://arxiv.org/abs/2205.01423

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID214174 (September)