

Contents

Motivation and objectives	2 3 7
The ICOS Architecture	
arly Adopters: the project's Pilot Cases 7	
Roadmap	9
Who we are	10

0

This project has received funding from the European Union's HORIZON research and innovation programme under grant agreement No 101070177.



0



Motivation and objectives

A continuum, today also referred to as Cloud continuum, IoT continuum, Edge-to-Cloud or Fog-to-Cloud, is expected to provide the means for workloads execution and data processing both in the Edge and the Cloud. According to International Data Corporation's (IDC) forecast, the total Edge spending in the EU will reach 75 billion dollars by 2026. In Europe, Edge computing is expected to increase at an annual growth rate of 26,5%, from €815m in 2020 to €2.6bn in 2025^{1,2}. Ensuring the security of data at the Edge as well as setting-up and maintaining an Edge infrastructure that can support highly demanding workloads are among the main barriers to the Edge adoptions in the enterprises³.

Despite the intensive research activities and clear industrial trends in this field, performance and efficiency of resource usage as well as contextual intelligence of a continuum, remains a daunting challenge. This is not only due to the continuum being intrinsically heterogeneous, volatile, distributed and increasingly cognitive, but also due to the emerging request to be open and collaborative. Networking, AI, green computing and parallel processing are just some of the further research topics that need to be leveraged in order for the continuum to become mainstream. There is a real need for an integrated platform to unleash the potential of European providers across the continuum. Currently missing from the Edge-Cloud scape is an open, non-proprietary, interoperable, robust, secure, sustainable multi-Cloud and multi-Edge continuum hosting solution, aimed at optimising the execution of workloads, especially in data intensive applications, and able to adapt to different strategies (e.g., execution time reduction, concurrent execution, Edge processing, fog security, locality, high resource utilisation, low latency and high energy efficiency), while being scalable, extensible and open to experimentation.

ICOS continuum will contribute to an open ecosystem, enabling interoperability with existing and emerging frameworks, towards a collaborative European Edge market scenario. It will provide new opportunities to European actors to establish

The ICOS project will design, develop and validate a meta operating system for the continuum built upon the principles of openness, adaptability, data sharing and a future Edge market scenario, addressing the challenges of: i) devices volatility and heterogeneity, ii) continuum infrastructure virtualization and diverse network connectivity; iii) optimised and scalable

ICOS pushes the envelope towards the next generation of continuum management by proposing a high-level meta operating system (metaOS) to realise an extensible, open, secure, adaptable, AI-powered as well as well highly performant and technology agnostic continuum.

service execution t and performance, as well as resources consumptions, including power consumption; iv) guaranteed trust, security and privacy, and; v) reduction of integration costs and effective mitigation of Cloud provider lock-in effects.

1. Rowan, Brendan, Álvarez, José Enrique, & Kušíková, Zuzana. (2023). Technology scoping paper (1.0). Zenodo. https://doi.org/10.5281/zenodo.8020703

2. Worldwide Edge Spending Guide (2021). IDC. https://www.idc.com/getdoc.jsp?containerId=IDC_P39947 3. Future Enterprise Resiliency and Spending Survey (2022). IDC. https://www.idc.com/getdoc. jsp?containerId=US48925022



The ICOS Architecture

ICOS has been conceived as a dynamic metaOS distributed along the continuum. The major design principle in ICOS leverages the capabilities of both the Cloud (virtually unlimited computing and storage capacity, ubiquity) and the Edge (locality exploitation, latency and communication reduction, privacy preservation), to optimise both the usage of resources and the performance of the users' workloads. One of the most interesting characteristics of ICOS is its



Figure 1 ICOS Continuum nodes

elasticity, being able to manage dynamic and mobile nodes efficiently and seamlessly. Thus, nodes can join or leave the system dynamically, or move throughout the continuum, establishing new proximitybased relationships between other nodes in different geographic locations. To realise these scenarios, two types of nodes will compose the continuum (Figure 1): the Agents that execute the users' workloads, and the Controllers that are responsible for managing the agents. The architecture also includes a 'Lighthouse' functionality to simplify and automate the dynamic on-boarding, disconnection, re-connection and migration of ICOS Agents to ICOS Controllers.

ICOS Agents are distributed along the continuum, ranging from constrained computation devices at the far Edge to high performance computing resources at the Cloud level. IoT and storage devices are attached to Agents, so both data access and computation of users' workloads are executed on Agents. The ICOS metaOS has been designed to be able to provide a very large number of heterogeneous Agents distributed over a wide geographical area. In order to efficiently manage such a huge number of nodes while exploiting the advantages of locality, a set of ICOS Controllers should be deployed along the continuum to cover the whole geographical area. ICOS Controllers are distributed with a flat, unstructured organisation. ICOS will run both on Agents and Controllers realising a unique, distributed Continuum middleware.

The ICOS conceptual architecture is built on four functional layers (Figure 2): the Distributed Meta-Kernel Layer, the Intelligence Layer, the Security Layer and the Data Management Layer. All layers are distributed along the ICOS Continuum providing functionalities at node level (e.g., node security assessment) as well as at continuum level (e.g., global workloads deployment optimisation). This widely distributed approach allows to decentralise the management of the continuum, reduce data transfers, ensure privacy and better exploit computational resources at the Edge without the need for a central point of control. An additional module, referred to as ICOS Shell, includes the user interfaces and the tools to interact with the ICOS continuum.



Figure 2: The ICOS conceptual architecture

The **Distributed Meta-Kernel Layer** provides the base functionalities necessary to make Cloud and Edge resources manageable and ICOS-ready: the set-up and maintenance of an ICOS Continuum, and the execution of



user's workloads. Key modules of this layer are described next.

The **Continuum Manager** is responsible for: i) registering and configuring resources (Cloud resources and Edge devices) to the Cloud Continuum; ii) discovering, labelling, and classifying the Edge devices and the IoT devices; iii) enabling the infrastructure and operational aspects to fulfil the remote execution of the user's workloads on the ICOS Continuum, and; iv) monitoring the status of the continuum and enforcing system and application policies.

The Runtime Manager is responsible for ensuring the underlying devices to fulfil execution requests from the upper layers of the ICOS metaOS (i.e., the ICOS Shell). The Runtime Manager components will: a) convert the service execution request into a workflow of tasks and generate an execution strategy; b) distribute the execution of tasks into multiple devices and orchestrate their execution; c) enforce optimal allocation policies and react to anomalies adapting the execution, and; d) ensure that the user workload is performing according to the expected Service Level Agreement (SLA) criteria. These functionalities are realised thanks to a joint collaboration of three main components: the Job Manager, deployed in the ICOS Controllers and receiving execution requests; as well as the Distributed & Parallel Execution and the Workload Offloader both deployed in the ICOS Agents and controlling the actual execution.

The Logging and Telemetry facilitates the collection of telemetry from the whole continuum, providing vertical application operators with a single view of performance, errors, logs, and component availability. Its goal is to effectively run and operate an end-to-end, multi-tenant, easy-to-operate and scalable observability system on the ICOS metaOS. The component enables the ingestion, long-term storage, and use of common observability signals, such as Quality of Service (QoS) metrics of applications and resources, logging and tracing from the entire continuum under a single consistent system with well-defined tenancy APIs and signal correlation and prediction (using the functionalities of the Intelligence Layer) capabilities. Monitoring a large and diverse set of distributed computing resources and software, generates a large amount of data of various types and may require these data to

be analysed, pre-processed, and aggregated if necessary, and transmitted from Edge nodes / IoT devices where monitored services are deployed to locations where end-toend services and resources management decisions are made. To prevent resource and application monitoring data from further contributing to the data overload, intelligent mechanisms will be implemented by the Monitoring and Telemetry component to automatically and dynamically decide what data to collect (i.e., what type of monitoring data to measure), the granularity of the data (i.e., what level of information to collect for a given item), and the frequency of the data (i.e., the interval between two collections for a given item).

The **Security Layer** is responsible for guaranteeing the security of ICOS users, resources, and applications at all times. It includes modules for authentication and authorization operations in the system, assess security of resources and applications and suggest mitigation actions, proactive discovery of anomalous behaviours and verification of the compliance of resources and applications. The main modules of this layer are introduced next.

The layer is coordinated by the **Security Layer Coordination** module that provides a unified interface for interacting with the Security Layer functionalities. It also manages data flows inside the layer and provides additional logic needed for seamless interactions between Security Layer's modules (e.g., task scheduling and periodic tasks).

The **Anomaly Detection** module is responsible for detecting anomalies in system and application logs. The Anomaly Detection module follows two workflows. The first workflow is a typical log-based anomaly detection workflow, consisting of log template extraction, learning normality, and anomaly detection⁴. The second workflow is the time series models using the numerical features extracted earlier from log-templates.

The **Security Vulnerability Mitigation** module operates based on rule triggering by which fulfilment of a certain rule initiates the execution of a certain recovery or mitigation processes. The latter can either be fully independent and automated or require the user's input. The basic rules and responses are pre-defined in static maps that can be further modified by the end user and also the

4. Jan Antić, Joao Pita Costa, Aleš Černivec, Matija Cankar, Tomaž Martinčič, Aljaž Potočnik, Gorka Benguria Elguezabal, Nelly Leligou, & Ismael Torres Boigues. (2023, May 15). Runtime security monitoring by an interplay between rule matching and deep learning-based anomaly detection on logs. <u>https://doi.org/10.5281/zenodo.7937448</u>



Intelligence layer's AI Analytics module will offer further mitigation strategies and rules.

The **Audit module** is responsible for executing a series of light-audit checks, defining specific checkpoints that need to be passed to consider the audit successful. Those checkpoints explore different aspects of the Cloud Continuum realisation and take place upon the on-boarding of any infrastructure, ICOS element and resource. The goal of the component is to identify potential vulnerabilities and risks, either during on-boarding or when triggered during runtime by the Security Coordination Module. Furthermore, it can assist in developing a security compliance checklist that may be tailored to each deployed application.

The **Security Scan** actively checks for security issues within deployed ICOS resources (at runtime and potentially at the design time⁵), detects issues and recommends mitigations and/or recovery processes (with help from the Security Vulnerability Mitigation module). The module may be called by the Audit module in case a security audit is to be enforced.

The **Compliance Enforcement** module is responsible for defining (and, in general, managing) security compliance policies for resources and applications in the ICOS System, automating the verification of the active policies and triggering remediation activities for compliance violations. Compliance Enforcement will work in conjunction with the Security Vulnerability Mitigation module for the definition of remediation activities.

The **Identity and Access Management** module has the overall objective of ensuring that the right people will have access to the right resources in the system. The three main responsibilities of this module are, the management of the users and their roles/ permissions, the authentication of the users and the authorisation of the requests within the system. All ICOS components will rely on this module to make sure that the received requests are properly authenticated (the requesting entity is known, and it is confirmed that it is who it says to be) and authorised (the requesting entity has the requested privileges to do that request).

Trust and Privacy are fundamental principles on top of which the Security Layer, and the entire ICOS metaOS, is built. Trust will be implemented at the system-wide scale using a communication protocol that enables automated and assisted distributed systems to publish secure data which can be subscribed to by the safety monitoring ICOS system (both TLS and mutual TLS transport protocols). The safety monitoring system could run on the same virtual or physical ICOS controller (i.e., computer/device, a separate computer/device on the IoT network or a remote computer at the Edge or in the Cloud). Privacy comprehends a set of fundamental elements which can be utilised for transforming data, including encryption and anonymisation. The anonymisation is a part of the Data Management and Intelligence layers. In the case of the former, data stored might need to be anonymised to ensure compliance with the GDPR's specific component requirements. The Intelligence layer contains the data processing module, which aggregates different functions to anonymise data. Additionally, another part of this functionality is to ensure information encryption in communication and storage by using internal SSL/TLS certificates.

The **Intelligence Layer** has the objective of bolstering and enhancing the functions and efficacy of the security and metakernel layers. It is responsible for delivering capabilities to facilitate the training, testing, deployment, maintenance, and updating of analytical and machine learning models across data-intensive applications in the Edge/Cloud spectrum, while adhering to particular data and model usage policies, focusing on ensuring trustworthiness. Next, the key modules of the Intelligence Layer are presented.

The layer is coordinated by the **Intelligence Coordination** module that provides a unified interface for interacting with the Intelligence Layer functionalities. It also coordinates the internal modules to fulfil the incoming requests.

The Data Processing module aggregates libraries and frameworks to allow data processing and transformation at scale in various devices, ranging from Cloud to lowend devices and user applications using the ICOS metaOS. It allows data Access & Storage and preprocessing using the Data Management module (batch, stream, local, remote).

The **AI Analytics** module aggregates libraries and algorithms to train and forecast using state-of-the-art methods to achieve a smart

5. Matija Cankar, Nenad Petrović, Joao Pita Costa, Aleš Černivec, Jan Antić, Tomaž Martinčič, & Dejan Štepec. (2023, April 15). Security in DevSecOps: Applying Tools and Machine Learning to Verification and Monitoring Steps. ICPE '23 Companion: Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, April 2023 (ICPE). <u>https://doi.org/10.1145/3578245.3584943</u>



meta-kernel and support the project ML use cases.

This component comprises a selected stack of streamlined machine learning algorithms and pipelines tailored to train and test predictive and optimisation models. It encompasses deep learning, adaptive and static batch machine learning, traditional machine learning, transfer learning and reinforcement learning libraries optimised to function efficiently on constrained devices. It considers models for both structured (e.g.; time-series) and semi (nested objects) or unstructured data (Text, Images). Advanced techniques, such as transfer learning, are to be considered in the second version of the module. The AI Analytics module will enable the proliferation of Artificial Intelligence/Machine Learning (AI/ML) algorithms aiming to empower autonomous network management, heavily facilitating, and automating the orchestration of the network resources. Automatic management of AI/ML workflows and lifecycles can act as a significant subcomponent towards zero-touch automation.

The ICOS Model Repository offers a collection of pre-trained analytics and ML models that may be reused, updated, modified (e.g., transfer learning), and integrated to promote the deployment of novel AI approaches across the ICOS metaOS layers. It includes capabilities for training and compressing these models for being used in constrained devices. This repository will make it easier to store pre-trained models of different versions that can be pulled and integrated with the existing infrastructure. The management of all algorithms and libraries used in the various metaOS versions will be possible thanks to this repository, which also supports seamless collaboration, repeatability, and strong version control of the models.

The Trustworthy AI module addresses data privacy issues by using federated learning and data anonymization. Additionally, it provides functions to make models explainable and transparent, explaining what leads to a decision or a forecast.

Its goal is to provide specific algorithms to analyse the datasets and develop models while maintaining the utmost respect for privacy and trustworthiness policies. This can be done by allowing models to be trained in a federated learning fashion to ensure data

protection in datasets containing userspecific data and by providing explainable Al algorithms that give meaningful insights into the output of models to the different layers in ICOS. Therefore, this component can be organised two-fold: 1) to ensure trustable, secure & robust model training via federated learning techniques, and 2) to provide model explainability through a series of AI interpretability algorithms, to aid the decision-making process of the models while understanding inputs and outputs. Both explainable AI methods and the ability for models to be trained via federated learning will be made available to provide data security in datasets, including user-specific data. This will give the various ICOS layers confidence in the models' output.

The Data Management Layer enables the communication between the different layers abstracting over the distribution of the system. Its main functionality is to ensure that the required data is available in those devices where it is needed and at the time that it is needed, to efficiently support ICOS operations. This layer will abstract infrastructure and communication details, so that the rest of ICOS components can focus on their specific functionalities and remain agnostic of the dynamicity and heterogeneity of the infrastructure. In addition, unnecessary data transfers will be avoided to reduce network congestion and increase overall performance of the platform. This layer also ensures that all data transfers between devices, as well as data at rest, are confidential.

In addition to the four layers, the **ICOS Shell** layer will allow users to interact with ICOS (e.g., infrastructure providers to manage their infrastructures, system integrators to manage their applications). The ICOS Shell will consist of three main components: the command line interface (CLI), the graphical user interface (GUI) and the development and operations (DevOps) tools. The Shell will be able to integrate with other tools and pipelines through support for machine-readable input and output.



Early Adopters: the project's Pilot Cases

The ICOS project includes four pilot cases that play an important role in the project, summarised in two critical aspects. First, pilot cases will provide functional and nonfunctional requirements for the ICOS System and, second, they will be early adopters of the system providing essential feedback on how to improve it. At the same time, the adoption of ICOS will enhance the pilot cases with unique application deployment and runtime management features implemented by ICOS.

Agriculture Operational Robotic Platform



The Agriculture Operational Robotic Platform (AORP) is an agro robot that can execute different tasks and missions, like sowing and tending crops, removing weeds, monitoring crop development, and identifying threats. The platform moves autonomously through the

field, performing the assigned missions. The robotic platform consists of control and driving modules. In addition, it is equipped with interchangeable tools - a seeder and a sprayer. The AORP is equipped with cameras, sensors and Edge computational devices that can be connected to the Cloud directly, via the transport platform, or via farm connectivity.

The **challenges** expected to be addressed in this pilot by ICOS are:

• Delays in accessing data affecting the limitation of field robots operating speed.

• Edge to Cloud orchestration of applications according to processing, or time requirements, improving the coexistence of real-time processing and coordination with Cloud services.

 Challenges in connectivity in real conditions and continuous monitoring of device operation.

The **benefits** expected by introducing the ICOS metaOS will be the reduction of the latency in decision taking by combining the computing capabilities on Edge and Cloud, the improvement of the AI models used during the missions, increasing the overall system availability, and the introduction of predictive maintenance in the AORP.



Railway Structural Alert Monitoring system

Today, the railway monitoring process is quite archaic. For most railway operators the inspection is done through physical inspections by railway staff or with a special train

with sensors on its wheels which runs through the whole rail system. Unfortunately, with these methods, measurements are taken only once or twice a year; in the remaining months, nobody knows what happens, and moreover, there is no established procedure to evaluate the cost-effectiveness of the actions taken to address the identified rail line issues. That is where the IoT devices provided by **Worldsensing** can minimise the monitoring and maintenance costs and allow operators to monitor in real-time important aspects like: rail tracks levelling, tensions and slope, surrounding areas settlements and falling elements, catenaries maintenance, cyber processes monitoring, etc.

The **challenges** expected to be addressed in this pilot by ICOS are:

 Implement energy-efficient solutions for low-power IoT devices to guarantee safety operation monitoring in real time while ensuring a very long lifetime of the deployed technology in remote locations. Improve raw data transmission and balance processing



between the Edge and Cloud.

• Improve wireless networking protocols to achieve reliable system operation in remote locations while ensuring connectivity management for the whole ICOS continuum.

• Edge to Cloud orchestration of several applications according to complexity, processing, or time requirements while using the same devices deployed and improving

the coexistence of real-time processing and coordination with Cloud services.

The **benefits** expected in this pilot case by adopting the ICOS metaOS will be a reduction of latency in decision taking (because some decisions will be taken at Edge level), an increase in the overall availability of the system, and an improvement in the data security.

In-car Advanced Infotainment and Multimedia Management system



In-car services are crucial for establishing a positive business case for the development of connected and autonomous vehicles. Mobility is changing, making vehicles an integral part of the customer digital world (or life): as vehicles become increasingly tech forward, the experience will expand beyond the screen to delight all the senses. The rise of autonomous and connected cars will also generate new demand for in-car infotainment and entertainment. The automotive pilot will offer innovative media content and services focused on tourism to enhance the user experience while travelling in a car and getting to know and explore new places. In more detail, this pilot case will provide the user with an in-depth tailored experience about the place they are visiting, showing the tourist extra information and knowledge around that new area in an immersive way. The possibilities of this technology will be exploited not only for interaction between users inside the car but also for others remotely.

The **challenges** expected to be addressed in this pilot by ICOS are:

• Ensure seamless user experience by optimising the distribution of multimedia content and maintaining high levels of quality of service (QoS) and quality of experience (QoE) also in case of low connectivity.

• Provide secure multiuser communication and interaction infrastructure able to ensure privacy and security of shared data.

The **benefits** expected by the introduction of ICOS will be an optimised distribution of multimedia content so that high-quality reception can be achieved even in lowconnectivity situations while avoiding service interruptions. Energy optimization with consequent impact on cost and sustainability are further taken-away expected by the innovative solution.

Energy Management and Decision Support system



The SSEA pilot aims at providing an Energy Management and Decision Support System (EMDS), using the ICOS continuum with data collected from five smart homes. Each house will be equipped with smart technology which may include Micro-generation systems: PhotoVoltaics (PV) or Wind Turbines, Electric Vehicles (EV), Heat pumps, Home energy storage, and Smart meters. Real time energy consumption will be measured with the use of inductive power monitoring clamps (IoT devices), that will send the collected data to an Edge device installed in the house managed by the ICOS metaOS.

The energy management system will generate personalised and optimised energy suggestions tailored to customer needs based



on AI models and sustainable solutions. The AI will dictate when/how energy will be used/ produced/stored, with hyperparameters adapting or updating through reinforcement and/or federate learning. The customers should be able to track improvements between actual vs optimised costs to build their trust in the energy management system.

The **challenges** expected to be addressed in this pilot by ICOS are:

• Provision of secure solutions, where data protection and data security must be ensured throughout all stages, including data collection, analysis, storage, and processing.

• Provision of customised, innovative solutions for optimal energy usage, and increase of self-consumption to pave the path towards households net zero emissions.

• Ensure viable and sustainable real time solutions in all settings including areas of poor connectivity.

The **benefits** expected by the introduction of ICOS will be the delivery of secure and efficient energy management systems based on advanced and reliable Machine Learning techniques for energy forecasting and hometo-home parameters sharing to avail of learnings obtained in other houses. Thanks to ICOS, the application will be able to leverage Cloud and Edge capabilities for real time solutions, with latency reduction, increased security and flexibility to tailor to customers' specific needs, increasing client satisfaction and retention.

Roadmap

The ICOS project started in September 2022 and will end in August 2025. The project follows an iterative development approach that includes analysis, design, implementation, integration and validation phases for each iteration. The project just concluded the first analysis phase and defined the **first version of the ICOS Architecture in May 2023** that is presented in this white paper. The fully detailed architecture is documented in the project's deliverable D2.2⁶.

In **May 2023** the project started the first iteration for the design and **implementation** of the overall system. The project aims at implementing a first set of integrated functionalities that can be validated and prove the value of the ICOS metaOS.

In **September 202**3 the project will launch its **first Open Call** to select 5 solution development projects to enrich the capabilities and functionalities of the ICOS metaOS. ICOS allocated IM€ budget for these new projects. The selected projects will implement their solutions in the course of 2024.

In **November 2023** the first release of the ICOS metaOS (**alpha version**) is expected. It will consist of a deployable and working system

that will be validated in the project's early adopters pilot cases.

In **February 2024** the first **validation** report from the pilot cases will be available. This will be an important milestone for ICOS since it will provide feedback that can help the project to improve and enrich the functionalities of the system.

From March 2024 the project will start a **new development iteration** that will bring to the release of two additional refined and enhanced ICOS metaOS releases: the beta version (June 2024) and the final version (April 2025).

In June 2024 the project will launch the **second Open Call** to select 15 ICOS uptake projects to validate the ICOS solution across various domains with a total budget of 900K€. The selected projects will use and validate the ICOS metaOS and will receive support to integrate ICOS in their solutions.

ICOS is part of the **EUCloudEdgeIoT.eu**⁷ initiative that promotes the collaboration between research projects, developers and suppliers, business users and potential adopters to realise a **common roadmap** for the understanding, the development and the adoption of the Cloud, Edge and IoT (CEI) Continuum in Europe. ICOS is active and will contribute mainly to the Open-Source Management, Architecture, and Market & Sectors task forces.

6. D2.2 - ICOS architectural design IT-1 (2023). ICOS project - <u>https://www.icos-project.eu/deliverables</u> 7. https://eucloudedgeiot.eu/about/

Who we are

ICOS is a project funded by the European Union's HORIZON research and innovation programme under the grant agreement N° 101070177. ICOS brings together 22 organisations (7 industry partners, 8 academia and research centres, 6 SMEs, and 1 organisation from the public sector) from 10 European Union countries, Switzerland and Israel. Background and experience of the partners covers different areas and

 \bigcirc

disciplines such as IoT-Edge-Cloud continuum management, AI, Security, OS development, Data Management, Networking and Ethics. This allows a multidisciplinary approach and creates a strong and solid basis for ensuring success of the project's activities.



icos_project @icos_project

 \mathbb{X}

D

This project has received funding from the European Union's HORIZON research and innovation programme under grant agreement No 101070177.

