AI-empowered Edge Cloud Continuum for self-aware cognitive computing environments

HORIZON Innovation Actions - HORIZON-CL4-2022-DATA-01-02

D2.1: Use cases, requirements, design and specifications

Delivery date: 06/2023

Dissemination level: PU

Project title:	**AI-empowered Edge Cloud Continuum for self-aware cognitive computing environments**
Duration:	**1 January 2023 – 31 December 2025**
Project website:	**https://www.cognifog.eu/**

Funded by the European Union under the Grant Agreement No. 101092968. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

Document information

Deliverable	D2.1: Use cases, requirements, design and specifications
Work Package	WP2: Requirements Definition and Architecture Design
Task(s)	T2.1: Use cases definition, requirements extraction
Type	R
Due Date	M06, 06.2023
Submission Date	30.06.2023
Document Lead	LMS
Contributor(s)	THALES, TMA, LMS
Internal Review	KENT, eBOS

Document history

Version	Date	Changes	Contributor(s)
V0.1	03/03/2023	Initial deliverable structure	LMS
V0.2	24/03/2023	40% of the deliverable content	THALES, TMA, LMS
V0.3	19/05/2023	80% of the deliverable content	THALES, TMA, LMS
V0.4	2/06/2023	Full draft	THALES, TMA, LMS
V1.0	13/06/2023	Internal Review Version	LMS
V1.1.1	19/06/2023	1st version with suggested revisions	KENT
V1.1.2	19/06/2023	2nd version with suggested revisions	eBOS
V1.4	26/06/2023	First revisions after internal review	THALES, TMA, LMS
V1.5	27/06/2023	Final Version for Quality Review	LMS
V1.6	27/06/2023	Revisions after final review	eBOS
V2.0	28/06/2023	Final version for submission	LMS

Consortium

Table 1: COGNIFOG consortium

Logo	Partner	Country	Short name
	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES FR	France	CEA
	ATOS IT SOLUTIONS AND SERVICES IBERIA SL ES	Spain	ATOS
	CYSEC SA CH Associated	Switzerland	CYSEC
	EBOS TECHNOLOGIES LIMITED CY	Cyprus	EBOS
	NETCOMPANY-INTRASOFT SA LU	Luxembourg	INTRA
	FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA ES	Spain	I2CAT
	KENTYOU FR	France	KENTYOU
	LABORATORY FOR MANUFACTURING SYSTEMS AND AUTOMATION	Greece	LMS
	TELEMATIC MEDICAL APPLICATIONS EMPORIA KAI ANAPTIXI PROIONTON TILIATRIKIS MONOPROSOPIKI ETAIRIA PERIORISMENIS EYTHINIS	Greece	TMA
	THALES France	France	THALES

SIEMENS	SIEMENS AKTIENGESELLSCHAFT DE	Germany	SIEMENS
UTA Urban Technology Alliance	Urban Technology Alliance Association CH Associated	Switzerland	UTA

List of abbreviations

Abbreviation	Definition
IoT	Internet of Things
WP	Work Package
TCP/IP	Transmission Control Protocol/ Internet Protocol
Gbps	Gigabit per second
LTE/5G	Long-term evolution/ 5th generation
KPI	Key Performance Indicator
USB	Universal Serial Bus
SWaP	Size, weight and power
OODA loop	observe-orient-decide-act loop
ISO	International Organization for Standardization
PACS	Picture Archiving and Communication System
RIS	Radiology Information System
EMR	Electronic Medical Records
MCU	Multipoint Control Unit
IT	Information Technology
JWT	JSON Web Tokens
OAuth	Open Authorization
API	Application Programming Interface
ECG	Electrocardiogram

ORM	Object-Relational Mapping
CI/CD	Continuous integration/ Continuous delivery
CEU	Central European University
MRP	Material Requirement Planning
AR	Augmented Reality
ROS	Robot Operating System
DT	Digital Twin
CPU	Central Processing Unit
HW	Hardware
SW	Software

Executive summary

In the context of WP2 and Task 2.1, the objective of this task was to identify use-cases reflecting stakeholder needs. With that in mind, the main goal was to choose or formulate three use case Trials that would highlight in the most efficient manner the COGNIFOG assets, objectives and features.

D2.1, as all the deliverables of WP2, will be living documents throughout the whole project duration. More specific, in D2.1 the Trials initial description is addressed. The document follows a discrete structure for all the use cases with the following sub-sections for each Trial:

- General description of company hosting each Trial
- Trial candidate scenarios
- Scenarios evaluation
- Technical description of the selected scenario
- Challenges and identified gaps

Three Trials are going to implement the COGNIFOG solutions. Each one of them introduces different challenges that need to be addressed with the COGNIFOG solution. For Trial 1 dynamic deployment following the workload requirements, Intermittent Connectivity and real time constraints in distributed systems are the main challenges that need to be tackled. For Trial 2 integration and deployment of new medical devices on telemedicine suitcase and updates, privacy issues, communication protocol support and intermittent Connectivity are the main pain points that with the support of COGNIFOG will be addressed. For Trial 3 the main challenges, as they are addressed in the initial stages of the project, are hardware devices communication methods, data coordination for Robot Operating System, security realization and integration and computing approaches optimization. To address these issues the KPIs that need to be enhanced for each Trial should be addressed. Moreover, in order to come up with a more applicable architecture and solution in general, the Trials potential scenarios (use case characteristics and requirements) were rated based on the COGNIFOG technologies that was possible to be identified in the project first six-month period.

As D2.1 stands as the first version of the use case requirements, design and specification, most of the inputs included in this document are going to serve as the base for other WPs and tasks as well. In the next version (D2.2) a more concrete description of the use cases assets and requirements will be detailed, as well as a rating of the COGNIFOG technologies application in each use case.

Table of contents

List of Figures

List of Tables

1 Introduction

In the scope of WP2 and Task 2.1, the objective of this task was to identify use-cases that align with stakeholder needs. With that in mind, the main goal was to choose or formulate three use case Trials that effectively showcase the COGNIFOG assets, objectives and features. To achieve this, dedicated workshops were conducted for the Trials where the use case scenarios were described, hardware and software components supporting the processes, as well as the potential challenges were discussed, particularly regarding the integration of the COGNIFOG technologies and methods.

To ensure ongoing engagement and involvement of the partners throughout the project activities, regular meetings were organized to discuss the progress of the WP2 activities, address any potential issues and devise solutions. In the first months of the project, the primary focus was on identifying potential application of the COGNIFOG technologies for the Trials scenarios. This involved continuous updates in D2.1 reflecting the advancements made in regular meetings and workshop.

The outcomes of T2.1 and the conducted activities carried out during the period, have resulted in addressing the use case requirements, which will serve as foundational initiatives for the subsequent tasks. This document encompasses the Trials definitions, requirements and challenges of the Trials, providing a baseline for the upcoming activities within the COGNIFOG project.

2 Trial 1 - Collaborative missions in urban areas

Thales is a multinational company headquartered in France, with operations in over 50 countries worldwide. The company specializes in the design and manufacture of high-technology solutions for customers in the aerospace, defence, transportation, and security industries.

Their activities include the development of communication systems, radar and surveillance technologies, navigation systems, and cybersecurity solutions. Thales is also involved in the design and manufacturing of satellites, avionics systems.

As of 2022, Thales has approximately 81,000 employees worldwide and reported revenue of €17.5 billion and a budget of 675 invest in research and development. With over 80 years of experience, Thales has a reputation for delivering innovative and reliable solutions to complex challenges, serving customers ranging from governments to commercial enterprises.

2.1 Trial Candidate Scenarios

For the THALES Trial, two candidate scenarios have been formulated and described in the following sections.

2.1.1 Scenario 1 - Detection of inundation and disaster-affected zones

In order to detect an inundation zone, a series of interconnected steps are undertaken to gather, analyse, process, and transmit data for accurate and real-time flood detection and monitoring.

The process begins with data collection for detection of inundation zone from sensors including water level sensors, LiDAR, barometers. Cameras, gyroscopes, magnetometers, and accelerometers are used to detect if a building is about to collapse or if there are some infrastructure failures. The data is transmitted to the edge of the network wirelessly or through a wired connection.

Then, collected data from the sensors is processed at the edge of the network to filter and extract useful information. The edge data processing helps to keep the context awareness, such as the location, time, and environmental conditions, which is useful for flood and disaster detection. By preserving this contextual information, flood detection accuracy is significantly relevant.

Once the edge has finished its processing, the result is transmitted to the cloud where advanced analysis techniques can be applied on a larger scale. The global data is used to build a comprehensive picture of the flood zone and the extent of the flooding. After the data has been collected, analysed and processed. Visualization techniques are employed to represent the information in a meaningful and easily interpretable manner.

Overall, the detection of an inundation zone involves collecting and analysing data from multiple sensors and processing the data at the edge of the network before transmitting it to the cloud for further analysis. This allows for the real-time detection and monitoring of flood zones, which is essential for effective flood management and response.

2.1.2 Scenario 2 - Mission coordination

In order to plan people extraction and the delivery of basic necessities, the result obtained from the previous scenario will be utilized. Furthermore, drones equipped with cameras and sensors will be launched to detect individuals in need of rescue. These drones will be deployed using a cloud-based application.

Additionally, cloud-based applications and IoT resources will be deployed to coordinate rescue efforts and provide essential supplies. This involves utilizing the cloud to store and share information, such as maps, location data, and communication channels.

The rescue team will coordinate their activities on the field using mobile devices such as phones and tablets. Through these devices, they will share vital information, such as location data, status updates, and emergency alerts.

In summary, this mission coordination scenario utilizes a combination of drones with advanced sensors, including cameras, magnetometers, and accelerometers, to locate and rescue people during a flood. The rescue team will leverage cloud-based applications and IoT resources to coordinate their efforts, communicate with each other, and deliver essential supplies such as food, water, and medical equipment to those affected by the flood.

2.2 Scenarios evaluation and selection

To evaluate the potential Trial 1 scenarios, a workshop was organized. In this workshop, a more detailed description of the use case scenarios was presented from THALES, and all the technology providers shared their opinions for the possible application of COGNIFOG assets to the use case.

Following the discussion, the partners unanimously concluded that selecting scenario 2 for Trial 1 would effectively showcase the COGNIFOG objectives. In Table 2 the outcome of the workshop, regarding the evaluation of the Trial 1 scenarios based on the COGNIFOG assets is presented.

Table 2: Trial 1 Scenarios evaluation

Technology/Tool	Partner	Objective	Application	
			Scenario 1	Scenario 2
Multi-cluster deployment and run-time optimization	ATOS	O2, O4	++	++
Performance, Efficiency and Predictive-Cognitive Monitoring	ATOS	O2, O3, O4	+	+
Runtime self-adaptation policies enforcement	ATOS	O2, O3, O4	+	++
Framework for industrial grade DevOps	SIEMENS	O1, O2,	+	+
Polygraph	CEA	O1, O2, O3	++	+
XanthOS	CEA	O1, O4	--	--
Intelligent RAN Controller	i2CAT	O1, O4	++	++
Slice Manager	i2CAT	O1, O4	+	+

IoT Edge Gateway	i2CAT	O1, O4	+	+
CYSEC Lab : Security Analysis and Risk Assessment	CYSEC	O1, O4	+	+
CYSEC Crypto Service (KMS)	CYSEC	O2, O4	+	+
CYSEC ARCA Trusted OS	CYSEC	O2, O4	+	++
Application layer front-end components (Dashboard & APIs)	EBOS	O4	++	++
Data Platform - sensiNact Data Hub	KENT	O1, O4	+	+
Digital Twin – Composition of IoT microservices	KENT	O1, O4	-	++
Data Hypervisor – Kentyou Eye	KENT	O1, O4	+	++

++ Matching perfectly

+ Applicable

- Applicable with some interpretation

-- Not applicable

2.3 Technical description of the selected scenario

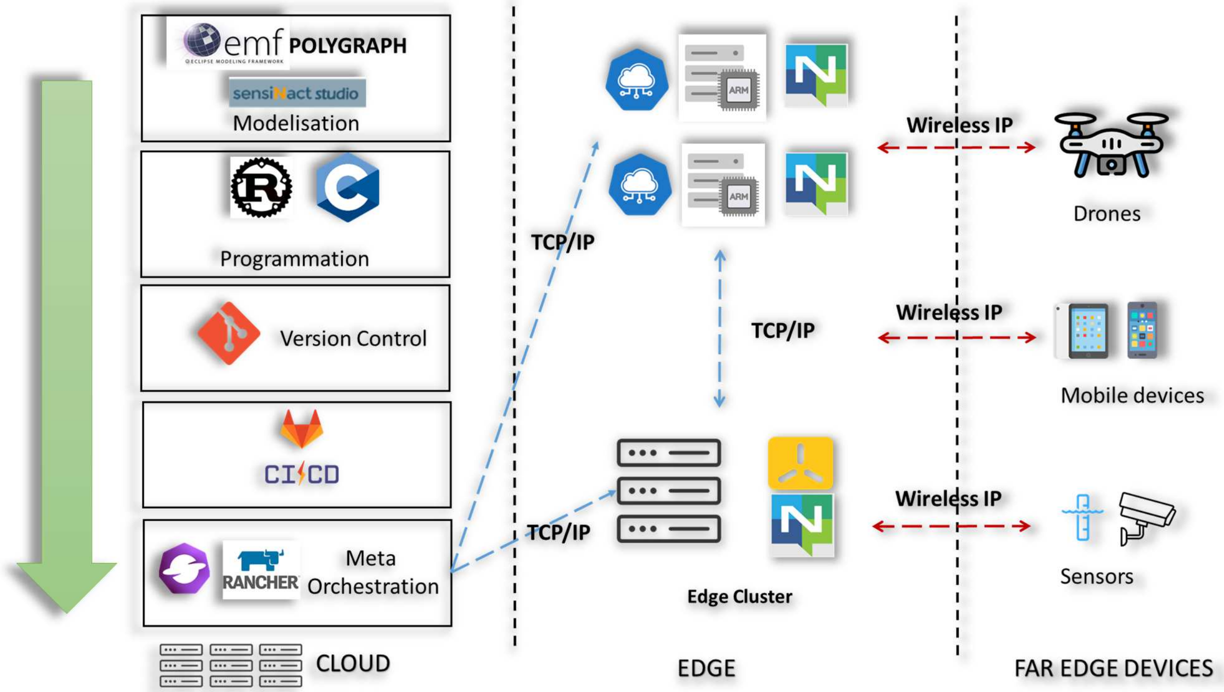

Figure 1: Trial 1 use case architecture

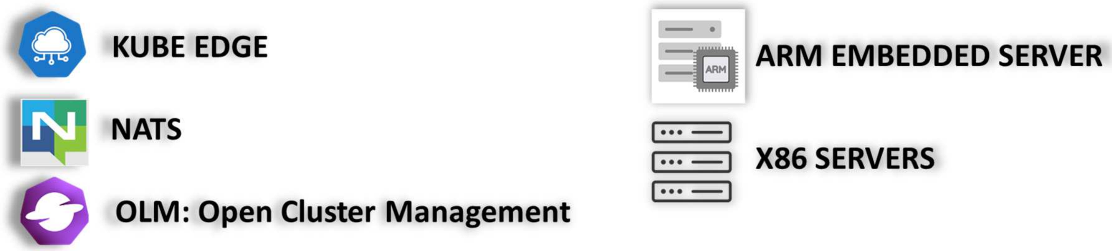

Figure 2: Trial 1 SW/HW components

Communication protocols:

- Communication from the cloud to the edge is typically carried out using TCP/IP protocols, which ensure reliable data transmission. This can be achieved through landlines, which can provide high bandwidth (up to 10 Gbps), or through wireless communications for vehicles, utilizing LTE/5G networks.
- Edge-to-edge communication primarily relies on wireless communication. The available bandwidth for such communication is typically below 1 Gbps and may be intermittent in nature.
- Far edge devices, such as sensors, drones, or mobile devices, utilize wireless communication with IP protocols to communicate with edge servers. The bandwidth for this communication is generally below 1 Gbps and can be intermittent.

Hardware:

- The cloud infrastructure operates on standard x86 servers deployed in both private and public clouds.
- Edge servers are typically equipped with x86 processors and are located in local fire and police stations. In addition, intervention vehicles utilize embedded boards with arm64 processors for their edge computing requirements.
- Far edge devices are powered by microcontrollers that incorporate sensors and wireless connectivity. Additionally, certain far edge devices may employ embedded boards (ARM64) with cameras and specialized accelerators, such as neural processing units, to enhance their capabilities.

In Table 3 the current Trial KPIs are presented. A more detailed version is presented in D2.5: Validation and benchmarking plan initial version.

Table 3: Trial 1 KPIs overview

KPI	Description
Application deployment time on Edge and IoT	Currently all is done manually with physical media like USB key or hard drives so it required human intervention
Deploy at least 100 micro applications that will be monitored by COGNIFOG platform tools with using 10 variability level domains	In cloud environment, the hardware is standardized and homogeneous (ex: x86 Server Rack)
Perform large scale monitoring and data collection from multiples devices	Currently we are able to monitor large scale deployment at cloud or at the edge but independently. There is no means to collect essentials information from far edge to cloud with SWAP constraints.
Perform large scale monitoring of OODA loops of actuators.	Currently we are able to monitor large scale deployment at cloud or at the edge but independently. There is no means to collect essentials information from far edge to cloud with SWAP constraints.
Process and aggregate the data at the edge to minimize the Edge to cloud bandwidth.	Currently we are able to monitor large scale deployment at cloud or at the edge but independently. There is no means to collect essentials information from far edge to cloud with SWAP constraints.
Optimize the edge computing process to minimize the edge computing power needed at edge	Currently we are able to monitor large scale deployment at cloud or at the edge but independently. There is no means to collect essentials information from far edge to cloud with SWAP constraints.
Satisfy and monitor 10 hard real-time classes (Hard real-time means that tasks have to be executed within a maximal time constraint)	Currently there is no global reference time for distributed systems (example: cloud to edge/ edge to edge). So it is difficult to meet time constraints.

2.4 Challenges and identified gaps

Trial 1 challenges overview is presented in Table 4 and a more detailed description is provided below:

<u>Dynamic deployment following the workload requirements:</u>

The problematic of dynamic deployment in continuum computing revolves around achieving efficient, adaptable, and energy-efficient resource allocation based on evolving workload requirements in a heterogeneous and mobile environment. Firstly, the problematic of dynamic deployment in continuum computing arises due to heterogeneity. Continuum computing environments consist of devices with varying capabilities, such as processing power, memory, energy constraints, and network connectivity. Deploying applications dynamically requires considering the characteristics of each device to ensure efficient resource allocation. Secondly, workload variability is a significant challenge in continuum computing. Workloads can change over time in terms of computational requirements, data processing needs, and communication patterns. Dynamic deployment must adapt to these changes to optimize resource utilization. Thirdly, mobility introduces additional complexities to dynamic deployment. Users and devices in continuum computing are often mobile, transitioning between different environments and network conditions. Applications need to seamlessly switch between devices while maintaining their state and performance. Fourthly, energy efficiency is a crucial concern in continuum computing. Resource-constrained devices, such as smartphones and edge devices, require energy-efficient resource allocation strategies. Dynamic deployment should optimize overall energy consumption while meeting the performance requirements of applications. Addressing these challenges necessitates intelligent resource management and allocation mechanisms. Techniques such as workload prediction and modelling, context-awareness, adaptive resource provisioning, and efficient load balancing algorithms are employed. Effective communication and coordination between devices and cloud servers are also essential for seamless application migration and state synchronization.

Intermittent connectivity:

Intermittent connectivity in continuum computing refers to the challenges posed by sporadic or unreliable network connections within the computing environment. It can be attributed to factors such as network instability, mobility, and the limitations of edge environments. Firstly, network instability plays a significant role in intermittent connectivity. Factors like signal interference, network congestion, and varying signal strengths can lead to frequent disruptions and unreliable connections among devices. Secondly, mobility of users and devices within continuum computing environments contributes to intermittent connectivity. As devices move, they may encounter areas with weak or no network coverage, causing temporary disconnections or limited connectivity until a stable network connection is restored.

Real time constraints in distributed systems:

Real-time constraints in distributed systems within continuum computing refer to the need for timely and predictable responses to events and tasks. Real-time applications require guaranteed response times within specific deadlines, ensuring that critical operations are completed within the desired timeframe. In the context of continuum computing, where applications span multiple devices and environments, ensuring real-time constraints becomes even more challenging.

Table 4: Trial 1 challenges overview

Challenge	Description
Dynamic deployment following the workload requirements	Achieving efficient, adaptable, and energy-efficient resource allocation based on evolving workload requirements in a heterogeneous and mobile environment
Intermittent connectivity	Challenges posed by sporadic or unreliable network connections within the computing environment
Real time constraints in distributed systems	Need for timely and predictable responses to events and tasks

3 Trial 2 - E-health services in the Edge-Cloud Continuum

3.1 Company description

Telematic Medical Applications (TMA) is a leading healthcare systems integrator and value-added solutions provider in the area of computer science-based healthcare systems.

TMA studies, designs, supplies, installs, supports and maintains integrated solutions certified according to the ISO 9001:2008 standards. TMA is also certified with ISO 13485:2003 & ISO 1348: 2004, for the distribution of medical-technological equipment. This ensures that all processes undergo essential control on issues such as confidentiality, integrity and availability of information, to protect the medical data and involved resources.

Telematic Medical Applications is growing rapidly. Its main activities include designing, manufacturing and implementation of Healthcare and Telemedicine integrated systems, audio and video conference solutions, telematics, as well as projects of structured wiring, computer rooms and infrastructure. TMA is experienced with solutions and services in the following business sectors:

- Clinical telemedicine and complete management of all radiological and introspection depictions as well as biosignals-telemedicine and tele-radiology products.
- Integrated informative systems (PACS – RIS) and teleradiology.
- Ophthalmological medical record integrated informative systems (EMR).
- Homecare products. Homecare, telecare and e-Health products.
- Medical equipment.
- Videoconference Systems -Videoconference Multi Point Units (MCU).
- Multimedia Equipment for each type of installation - Audio-visual equipment for every type of installation (personal, meeting rooms, conference halls etc.).
- Integrated data networks (wired, wireless, satellite).
- IT and network safety systems.
- Electromechanical and network infrastructures - data centres

3.2 Trial Scenario: A flexible approach to e-health services running on the edge with federated orchestration

To make the scenario description easier to follow, the following naming is used: ***Station****,* ***Backend****,* ***Hub****. Better explanation of each resource is given in the technical description section.*

Currently all **Stations** (the telemedicine suitcase e-health systems) connect to a centralized server, where accounts are administered from patient level to municipality level. Although this is very practical, it does not scale very well.

However, a bigger issue are requirements of health providers and privacy concerns: Municipality-level customers may not want their resident's data (that can contain medical data) stored on a central server, but on-premises, hosted on physical machines within their jurisdiction.

In this scenario, one or more edge servers in the premises of a local medical call centre run the **Backend**. Administration however of the telemedicine devices and the multiple edge servers should be orchestrated by a **Hub**, based on customer agreements.

COGNIFOG components could help solve the following challenges:

1) Simplify privacy and scalability issues by running the Backend on edge servers.

If edge servers connect doctors with patients at the edge of their jurisdictions or area of interest, problems with privacy (patient cases would stay locally) and scalability can be solved.

2) Low latency and performance predictability

Stations can benefit by connecting to edge servers. Medical personnel may also login to the **Backend** on-premises or via internet to evaluate patient data from **Stations** and provide teleconferencing services.

3) Security issues

Limiting data exchange between different platforms and layers improves security. However, in a real-use scenario, we would need not only the **Stations** to connect to the edge server, but also third-party mobile apps, using secure connections with JWT & OAuth2 bearer token authentication.

4) Self-adaptability to poor/intermittent network connectivity

Relating to connectivity between IoT medical devices and edge servers, but also between edge servers and cloud datacentre. Measurements made by the **Station** cannot be immediately relayed to the **Backend** if offline. It would be beneficial to improve connectivity under poor or intermittent connectivity, so remote medical staff can evaluate the patient's measurements as quickly as possible.

5) eHealth in Smart Cities

Provide cumulative statistics of our measurements, anonymization and presentations on various "Smart City" portals and apps.

Most of our customers are municipalities that would like to present their social contributions in the e-health sector. We envision a type of dashboard that could collect (via an API with our Hub) the measurements of our **Stations** (which can be considered an IoT device), plus other metadata (GPS coordinates for example).

After anonymization it could create presentation data that can be used in portals of smart city users to present statistics, heatmaps, cumulative results of the e-health solution.

3.3 Scenario evaluation

To evaluate the potential Trial 2 scenarios, a workshop was organized. In this workshop, a more detailed description of the use case scenarios was presented from TMA, and all the technology providers shared their opinions for the possible application of COGNIFOG assets to the use case.

After discussion, it was decided that the Trial 2 use case could be strengthened if all the scenarios were combined to one. Hence, for the TMA use case, an evaluation of the use case scenario was conducted from all partners. In Table 5 the outcome of the workshop, regarding the evaluation of the Trial 2 scenario based on the COGNIFOG assets is presented.

Table 5: Trial 2 Scenario evaluation

Technology/Tool	Partner	Objective	Application Scenario 1
Multi-cluster deployment and run-time optimization	ATOS	O2, O4	++
Performance, Efficiency and Predictive-Cognitive Monitoring	ATOS	O2, O3, O4	++
Runtime self-adaptation policies enforcement	ATOS	O2, O3, O4	+
Framework for industrial grade DevOps	SIEMENS	O1, O2,	++
Polygraph	CEA	O1, O2, O3	++
XanthOS	CEA	O1, O4	+
Intelligent RAN Controller	i2CAT	O1, O4	++
Slice Manager	i2CAT	O1, O4	++
IoT Edge Gateway	i2CAT	O1, O4	-
CYSEC Lab : Security Analysis and Risk Assessment	CYSEC	O1, O4	+
CYSEC Crypto Service (KMS)	CYSEC	O2, O4	+
CYSEC ARCA Trusted OS	CYSEC	O2, O4	++
Application layer front-end components (Dashboard & APIs)	EBOS	O4	++
Data Platform - sensiNact Data Hub	KENT	O1, O4	+
Digital Twin – Composition of IoT microservices	KENT	O1, O4	++
Data Hypervisor – Kentyou Eye	KENT	O1, O4	+

++ Matching perfectly + Applicable - Applicable with some interpretation -- Not applicable

3.4 Technical description of the selected scenario

3.4.1 Summary of infrastructure provided

1) Portable tele-medicine system [Station]

The product that is featured in the trial scenario and can be considered as an edge device, is a portable self-contained telemedicine system that can be configured with medical devices for accurate patient assessment in mobile or remote applications. Its form is a rugged suitcase and can manage several medical sensors such as ECG, blood oxygen, glucose etc. It also contains camera/microphone for video conferences with medical professionals.

Medical situations that require a quick and flexible approach to patient care, benefit from this type of mobile telemedicine system. TMA No.A.H. Telemedicine Case has instrumental use in healthcare applications such as disaster response, mobile medical units, correctional facilities, home health care, military field hospitals, oil rigs, cruise ships, emergency medical care, and remote healthcare clinics.

The **Station** communicates with "e-Pokratis" **Backend** and via a https channel using JWT & OAuth2 bearer token authentication.

Figure 3: TMA tele-medicine device

2) E-Pokratis telemedicine platform [Backend]

A web-based patient record system designed specifically for the needs of telemedicine consultations.

The service provides a complete and accurate summary of an individual's medical history which is accessible online by the patient & authorized (by the patient) medical and paramedical personnel.

Health data on the backend includes patient-reported outcome data, lab results, and data from integrated medical devices.

Our backend connects to the **Stations** and other mobile applications. The web application allows managing customers, patients and medical personnel. Patient cases and files can be created and shared with consulting physicians.

Another important feature is orchestrating real-time video conferences. Live communication with a doctor is provided whenever necessary through a simple video call with the attending physician.

Video demonstration of the call centre: https://youtu.be/rk8VmkprNe0

3) Orchestrator backend hub [Hub]

This backend will run on a central cloud and will be developed to provision and update the various edge servers running the **Backend** and provisioning and updating the **Stations.** This is required to manage customers using the hardware and software negotiated in SLAs and for verification of overall topology, connections and energy consumption.

3.4.2 Technology stack

- **Telemedicine station** using .NET Framework
- **Mobile app** using Xamarin.Forms
- **Backend service/Web application:**
    - .NET 7 Core
    - ASP.NET Core with Blazor server-side frontend
    - DevExpress ExpressApp frontend
    - Database currently MS SQL, but database agnostic ORM has been used.

3.4.3 Server requirements

The back end can be compiled for any operating system (Linux/Windows) or processor and can be containerized. For Linux, the required packages are:

- aspnetcore-runtime-7.0
- libgdiplus

Standard dedicated user account for running service using supervisor and nginx server are used.

The service can work with any database (backend service is database-agnostic) like PostgreSQL (simplifies replication in a cluster) or MySQL.

Important note: The **backend** (if running on the edge) connects to the stations via websockets. The protocol is necessary for orchestration of teleconferences between the **Station** and **Backend**.

In Table 6 the current Trial KPIs are presented. A more detailed version is presented in D2.5: Validation and benchmarking plan initial version.

Table 6: Trial 2 KPIs overview

KPI	Description
Edge and IoT autonomous installation time (K1)	Currently there are no edge servers.
Setup time of infrastructure (K2)	Currently, updates of telemedicine devices are done manually via USB.
Service establishment time (K3)	In a simple client & central server environment there is no delay.
Single touch orchestration (K4)	Currently, this the approximate time to setup a new device.
Device integration (K6)	In a simple client & central server environment there is no complication.
Scale-up latency (K5)	Cloud software can work behind a load balancer, although in an edge model, requirements will be much less.
Data integrity / security (K8)	Moving data from cloud to edge provides better security. Encrypted backup however should be made from edge servers to cloud.
Number of test users (trial specific)	Emulation of 100 stations connecting with the backend, and maintaining websocket connections.
Number of test apps (trial specific)	Apps like mobile application connecting to backend, backend running on an edge server, automated backup application, third – party smart city dashboard, connectivity dashboard or CI/CD could be integrated

Reduce in latency of response (trial specific)	Stations connecting to edge servers is a new concept, however compared to connectivity with a central backend, the latency improvement is obvious.
Robustness (trial specific & K7)	The following communication tests should be used: 1) Station with edge server communication under (emulated) bad connectivity 2) Station with edge server under lack of any connectivity 3) Backend with Hub under lack of any connectivity Stress test: emulation of multiple stations communicating with edge server.

3.5 Challenges and identified gaps

To integrate new medical devices into the **Station,** testing is needed in a model station. If suitcases are deployed in various remote settings, updates should be reliable and automated. Moreover, provisioning of **Stations** based on SLAs is done manually, and becomes even more complex if the backend is running on an edge server. In addition, connecting suitcases to backends under various network conditions. The edge server running the **backend** and **suitcase** should be able to function even if connectivity with the central hub is lost. Lastly, to integrate medical AI for following self-diagnosis protocols, the edge server must be able to connect to the internet. Medical GPT services are still in a nascent level, but for the foreseeable future will only be accessible via a centralized API hosted in the cloud. An overview of the Trial 2 challenges is presented in Table 7.

Table 7: Trial 2 challenges

Challenge	Description
Integration of new medical devices on telemedicine suitcase and updates.	Any updates of the software running on the telemedicine devices is done manually and is extremely time-consuming.
Deployment	Provisioning of each telemedicine suitcase is done manually so integration testing can be performed on the medical devices.
Privacy concerns	All patient data is stored centrally in the cloud. This does not satisfy requirements for data localization.
Communication protocol support	WebSocket's should be supported by the connectivity layers. Required for orchestrating video-conference calls and other services.
Intermittent connectivity	Telemedicine stations often have very bad connectivity, or work offline. In this case measurements are stored and uploaded whenever internet is available.

4 Trial 3 - Automated Edge-Cloud continuum for smart manufacturing

4.1 Company description

Laboratory for Manufacturing Systems & Automation (LMS) is oriented toward research and development in cutting-edge scientific and technological fields. It belongs to the Mechanical Engineering and Aeronautics Department at the University of Patras. LMS is involved in research projects funded by the CEU and European industrial partners. Particular emphasis is given to cooperation with the European industry as well as with several "hi-tech" firms. It currently employs approximately 80 researchers organized into four different groups: a) Manufacturing Processes Modelling and Energy Efficiency, b) Robots, Automation, and Virtual Reality in Manufacturing, c) Manufacturing Systems, and d) Software development for manufacturing systems.

COGNIFOG will evaluate a brand-new paradigm for adaptable and flexible production starting from Trial 3, where LMS will act as the end user. By integrating flexible mobile dual arm robots controlled by flexible IT infrastructure that maximizes the exploitation of computer resources in multiple dimensions, such as throughput, prices, and energy savings, the goal is to build dynamically reconfigurable shopfloors. The use case for smart manufacturing is sufficiently wide to be used to a variety of manufacturing applications as well as to other uses, such warehouse management. MRPs are being implemented in many manufacturing industries, including the automotive, aerospace, white goods, and many others because they can support human operators throughout product assembly. LMS will use various methods to communicate the case to various industrial stakeholders.

Figure 4: LMS premises and infrastructure

4.2 Trial Scenarios

The goal of the use case for an Automated Edge-Cloud Continuum for smart manufacturing is to develop shop floors that can be dynamically reconfigured while maximizing flexibility and productivity. The major objective is to establish a dynamic and adaptable manufacturing environment that will allow manufacturers to easily adjust to evolving production needs. Industrial robots allow the system to quickly adapt to new product lines or production runs without a lot of retooling or manual work. However, there are several challenges to achieving this goal. One of the main challenges is to ensure that the advantages of industrial robots are fully exploited to optimize efficiency and productivity. Additionally, the system needs to reduce the latency of response or decrease the setup time of infrastructure. To investigate possible solutions and approaches toward a more effective industrial shop floor, two possible scenarios from the bus manufacturing line are proposed. In the following sections a brief description of the two candidate scenarios is presented.

4.2.1 Scenario 1 - Gluing of aluminium profiles

The process of aluminium profiles gluing has the following process steps and characteristics:

- humans pre-treat the profiles and panel areas and place the profiles to a working bench
- a robotic arm dispenses glue on the profile
- a vision system provides feedback for robot control and detects glue discontinuities on the profile
- human interacts with the workplace environment through AR glasses
- A scanner is watching the whole workplace to provide a level of safety by controlling the robot controller, since there is human robot collaboration.
- The process is supported by Simulation tools for process planning
- Digital Twin for virtual representation of the process

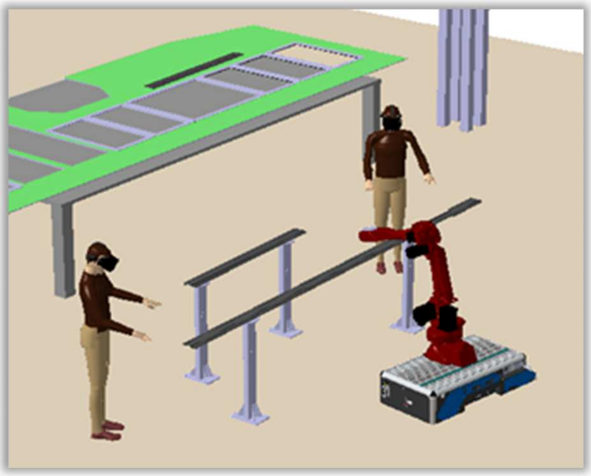

Figure 5: Scenario 1 graphical representation

4.2.2 Scenario 2 - Windows assembly on panel

The process of windows assembly has the following process steps and characteristics:

- A robot brings a window panel to the assembly area
- A robot picks up the window
- Glue is placed on the window panel
- Detect the corresponding installation point on the bus panel, through a vision system
- A robot places window on the bus
- The process is supported by Simulation tools for process planning

Figure 6: Scenario 2 graphical representation

4.3 Scenarios evaluation and selection

In order to select the most mature, flexible and applicable scenario based on the COGNIFOG assets, a workshop dedicated to Trial 3 was held. After the presentation of the scenarios, regarding the process and the assets involved in the use case, the main output of the workshop is shown in Table 8.

After discussions with all the partners and the evaluation of the scenarios, it was decided that scenario 1 was the most appropriate to be used for Trial 3 use case.

Table 8: Trial 3 scenarios evaluation

Technology/Tool	Partner	Objective	Application	
			Scenario 1	Scenario 2
Multi-cluster deployment and run-time optimization	ATOS	O2, O4	++	+
Performance, Efficiency and Predictive-Cognitive Monitoring	ATOS	O2, O3, O4	+	+
Runtime self-adaptation policies enforcement	ATOS	O2, O3, O4	++	-
Framework for industrial grade DevOps	SIEMENS	O1, O2,	-	-
Polygraph	CEA	O1, O2, O3	++	+
XanthOS	CEA	O1, O4	-	-
Intelligent RAN Controller	i2CAT	O1, O4	-	-
Slice Manager	i2CAT	O1, O4	-	-
IoT Edge Gateway	i2CAT	O1, O4	-	-
CYSEC Lab : Security Analysis and Risk Assessment	CYSEC	O1, O4	+	-
CYSEC Crypto Service (KMS)	CYSEC	O2, O4	+	+
CYSEC ARCA Trusted OS	CYSEC	O2, O4	+	-

Application layer front-end components (Dashboard & APIs)	EBOS	O4	++	++
Data Platform - sensiNact Data Hub	KENT	O1, O4	+	+
Digital Twin – Composition of IoT microservices	KENT	O1, O4	+	--
Data Hypervisor – Kentyou Eye	KENT	O1, O4	-	-

++ Matching perfectly

+ Applicable

- Applicable with some interpretation

-- Not applicable

4.4 Technical description of the selected scenario

As it was described in section 5.2, the use case scenario under investigation refers to the glue placement on aluminum profiles, with the establishment of a human robot collaboration. The overall architecture of the use case is presented in Figure 7. The operation is as follows:

- A human operator performs pre-treatment on the aluminum profiles (e.g., cleaning, degreasing) and places them on the working bench in front of the gluing dispenser.
- A robotic arm performs glue dispensing on the required aluminum profile areas guided by a multi-purpose vision system.
- After the gluing process, quality control is performed.

For the correct process realization, many components were utilized, as show in the architecture diagram. The operation of each one of the components is described below:

- **Robot Driver:** The software module that allows data exchange between ROS and robot controller.
- **Robot controller:** the robot controller is a hardware device that controls the robot operation.
- **Glue dispenser:** a device that enables dispensing of fluid/liquid materials.
- **Scanner:** a perception system that scans the workstation area and is used as a safety system
- **Camera:** a vision system that helps identifying potential gaps of glue on the profile and supports validation of the gluing operation
- **Human:** responsible for pre-treating the aluminum profiles and interacting with the robotized process through AR glasses
- **AR glasses:** wearable device that gives the ability to the human operator to interact with the whole system
- **Digital twin:** module that depicts the physical system within a virtual environment and enables data exchange for visualization, monitoring and interaction. . It also gives the ability for process tracking and performance evaluation to remote users/engineers.
- **ROS network:** Open-source framework - set of software libraries and tools. Among other features, ROS enables the communication of multiple devices under a common network structure.

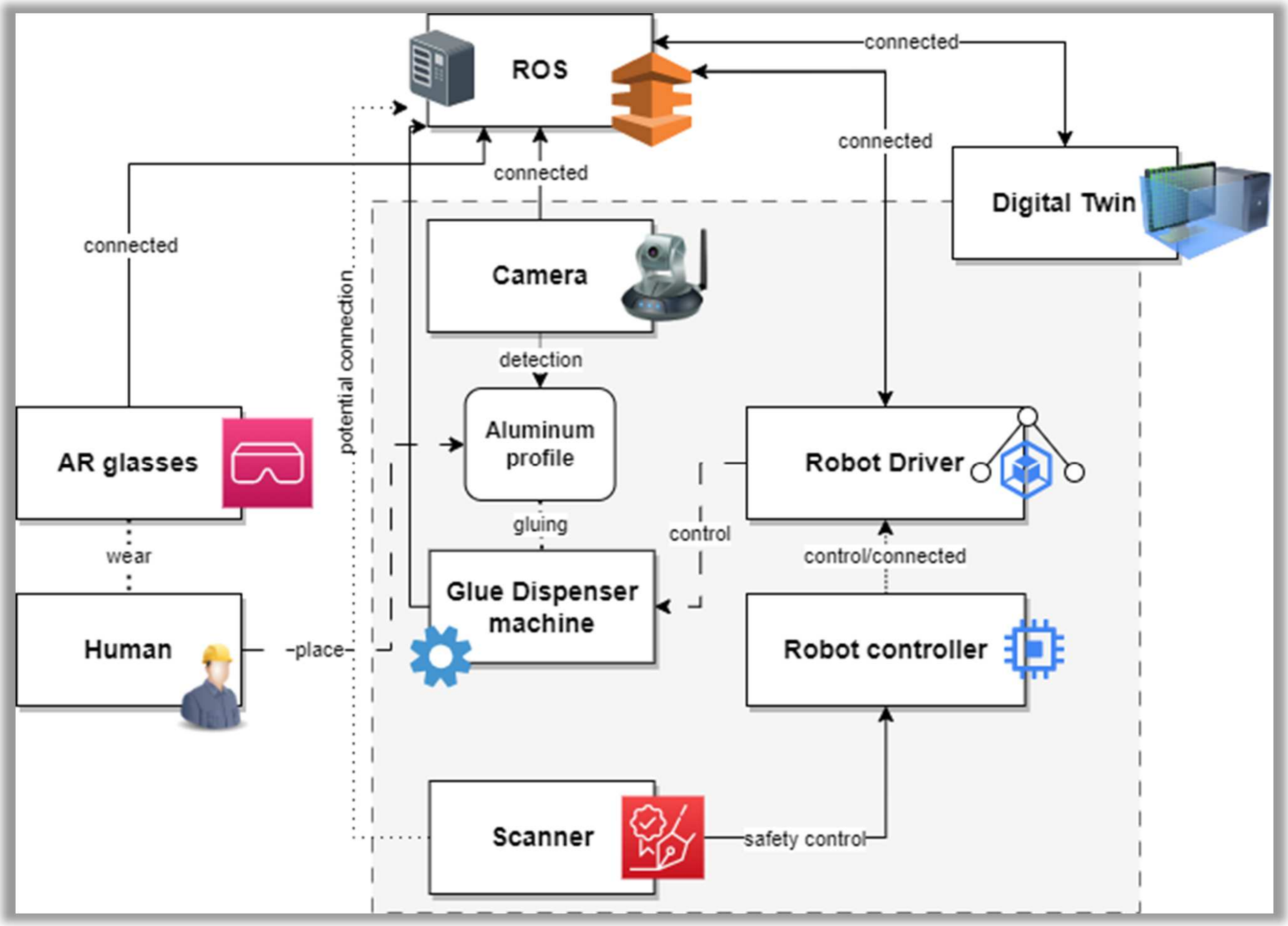

Figure 7: Trial 3 use case architecture overview

The technical characteristics of the use case assets (hardware and software) are presented in Table 9.

Table 9: Trial 3 HW/ SW components

Asset	Asset type	Description
Robot Driver	SW	ROS Interface for COMAU C5G controller
Robot controller	HW	COMAU C5G
Scanner/Data	HW/SW	Sick S3000
Camera/Data	HW/SW	ZED2 stereo camera
AR glasses	HW/SW	HoloLens2
DT	SW	Unity-based application
ROS	SW	Set of software libraries and tools

An overview of the Trial 3 use case KPIs are presented in Table 10. A more detailed version of the Trial KPIs is presented in D2.5: Validation and benchmarking plan initial version as well. However, during the rest of WP2 activities, more detailed KPIs will be addressed and they will be presented in the next version of this deliverable D2.2.

Table 10: Trial 3 KPIs overview

KPI	Description
Latency	Reduction of response times among the various use case components. Reduction of the time before scale-up operation is completed in case of a lack of resources

Setup time	Reduction of the elapsed time to setup a specific cloud-to-edge-to-IoT infrastructure
Service establishment time	Reduction of the elapsed time to setup a specific network service before it being operational
Security	To protect the network and sensor data from unauthorized access, manipulation, or disruption
Edge CPU utilization	Reduce the amount of Edge CPU usage by deploying some processes on Cloud

4.5 Challenges and identified gaps.

Trial 3 use case scenario introduces many challenges that could be addressed with the COGNIFOG technologies integration. Since the use case scenario offers the ability for multiple devices interaction and connection, the main issue is the efficient cooperation and control of all the hardware and software components. All the hardware devices that need to be coordinated, in order for the gluing operation to be completed in the context of human robot collaboration, need to utilize the ROS channels and data exchange methods. Another issue is the security provided for data and networks. If the whole coordination of the process would be a web application, then more security measures need to be implemented. However, when more devices need to be connected in a single network, then challenges such as lateness or insufficient data storage capacities need to be tackled. In Table 11 the main challenges identified for Trial 3 use case have been addressed.

Table 11: Trial 3 challenges overview

Challenge	Description
Communication of HW devices	Effective utilization of network capacity, Optimal exploitation of ROS capabilities
Coordination of data through ROS network	Bandwidth limitations extension, High frequency information exchange
Effective security integration	Security methods implementation
Optimization of existing computing approaches	Edge/Cloud computing techniques: Optimal distribution of data processing, Decentralized computing

5 Conclusion

Concluding, WP2 activities during this first period of COGNIFOG project had as main goal the initial definition of the use cases for each Trial. Developing a document (D2.1), that will be utilized as the baseline (along with D2.5) for all other activities, was the outcome of the T2.1 activities for addressing the specifications and characteristics of the Trials. Based on D2.1 (and D2.5), the architecture and the implementation specifications will be developed with a more concrete perspective, regarding each Trial unique requirements and characteristics. In WP2 next activities, a more detailed documentation of the Trials specifications (KPIs, requirements etc.) will be realized. In the next version (D2.2) a more technical and detailed description of the use cases will be provided, with the necessary adjustments and justifications if needed, after the other project activities take over to implement the COGNIFOG solution.