

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101069732



D2.1 – State-of-the-Art and market analysis report

Deliverable No.	D2.1	Due Date	30-NOV-2022
Type	Report	Dissemination Level	Public (PU)
Version	1.0	WP	WP1
Description	Document with main results of SotA review and stakeholders and market analysis comprehensive analysis.		



Copyright

Copyright © 2022 the aerOS Consortium. All rights reserved.

The aerOS consortium consists of the following 27 partners:

UNIVERSITAT POLITÈCNICA DE VALÈNCIA	ES
NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"	EL
ASOCIACION DE EMPRESAS TECNOLOGICAS INNOVALIA	ES
TTCONTROL GMBH	AT
TTTECH COMPUTERTECHNIK AG (<i>third linked party</i>)	AT
SIEMENS AKTIENGESELLSCHAFT	DE
FIWARE FOUNDATION EV	DE
TELEFONICA INVESTIGACION Y DESARROLLO SA	ES
COSMOT KINITE TILEPIKONINONIES AE	EL
EIGHT BELLS LTD	CY
INQBIT INNOVATIONS SRL	RO
FOGUS INNOVATIONS & SERVICES P.C.	EL
L.M. ERICSSON LIMITED	IE
SYSTEMS RESEARCH INSTITUTE OF THE POLISH ACADEMY OF SCIENCES IBS PAN	PL
ICTFICIAL OY	FI
INFOLYSIS P.C.	EL
PRODEVELOP SL	ES
EUROGATE CONTAINER TERMINAL LIMASSOL LIMITED	CY
TECHNOLOGIKO PANEPISTIMIO KYPROU	CY
DS TECH SRL	IT
GRUPO S 21SEC GESTION SA	ES
JOHN DEERE GMBH & CO. KG*JD	DE
CLOUDFERRO SP ZOO	PL
ELECTRUM SP ZOO	PL
POLITECNICO DI MILANO	IT
MADE SCARL	IT
NAVARRA DE SERVICIOS Y TECNOLOGIAS SA	ES
SWITZERLAND INNOVATION PARK BIEL/BIENNE AG	CH

Disclaimer

This document contains material, which is the copyright of certain aerOS consortium parties, and may not be reproduced or copied without permission. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The information contained in this document is the proprietary confidential information of the aerOS Consortium (including the Commission Services) and may not be disclosed except in accordance with the Consortium Agreement. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the Project Consortium as a whole nor a certain party of the Consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

The information in this document is subject to change without notice.

The content of this report reflects only the authors' view. The Directorate-General for Communications Networks, Content and Technology, Resources and Support, Administration and Finance (DG-CONNECT) is not responsible for any use that may be made of the information it contains.

Authors

Name	Partner	e-mail
Prof. Carlos E. Palau Ignacio Lacalle Úbeda Rafael Vaño Raúl San Julián	P01 UPV	cpalau@upv.es iglaub@upv.es ravagar2@upv.es aeros-project@outlook.com rausanga@upv.es
Dr. Harilaos Koumaras Vassilis Pitsilis Anastasios Gogos	P02 NCSR	koumaras@iit.demokritos.gr vpitsilis@dat.demokritos.gr angogos@iit.demokritos.gr
Eneko Rada	P03 INNO	erada@innovalia.org
Anna Ryabokon	P04 TTControl	anna.ryabokon@tttech.com
Philippe Buschmann José Eduardo Fontalvo Vivek Kulkarni Florian Gramss Korbinian Pfab	P05 Siemens	philippe.buschmann@siemens.com jose-eduardo.fontalvo-hernandez@siemens.com florian.gramss@siemens.com vivekkulkarni@siemens.com korbinian.pfab@siemens.com
José Ignacio Carretero Ken Zangelin	P06 FIWARE	joseignacio.carretero@fiware.org ken.zangelin@fiware.org
Ignacio Domínguez Diego López	P07 TID	ignacio.dominguezmartinez@telefonica.com diego.r.lopez@telefonica.com
Fotini Setaki George Limperopoulos	P08 COS	fsetaki@cosmote.gr glimperop@cosmote.gr
Kostas Giannoulakis Dimitra Siaili	P09 8BELLS	giannoul@8bellsresearch.com dimitra.siaili@8bellsresearch.com
Christos Xenakis Ilias Politis Panagiotis Bpountakas	P10 InQBit	chris@inqbit.io ilias.politis@inqbit.io Panagiotis.Bpountakas@inqbit.io
Dimitris Tsolkas Christos Milarokostas	P11 FOGUS	dtsolkas@fogus.gr milarokostas@fogus.gr
Joseph McNamara	P12 LMI	joseph.mcnamara@ericsson.com

Jimmy O'Meara		jimmy.o.meara@ericsson.com
Katarzyna Wasiliewska-Michniewska Maria Ganzha Marcin Paprzycki	P13 IBSPAN	katarzyna.wasielewska@ibspan.waw.pl maria.ganzha@ibspan.waw.pl marcin.paprzycki@ibspan.waw.pl
Tarik Taleb Tarik Benmerar	P14 ICT-FI	tarik.taleb@ictficial.com
Vaios Koumaras Vasileios Mavrikakis Nick Vrionis	P15 INFOLYSIS	vkoumaras@infolysis.gr vmavrikakis@infolysis.gr chsakkas@infolysis.gr nvrionis@infolysis.gr
Eduardo Garro M ^a Ángeles Burgos	P16 PRO	egarro@prodevelop.es mburgos@prodevelop.es
Michalis Michaelides Herodotos Herodotou	P18 CUT	michalis.michaelides@cut.ac.cy herodotos.herodotou@cut.ac.cy
Saverio Gravina Andrea Chentrens Sara Gaudino	P19 DST	saverio.gravina@dstech.it a.chentrens@dstech.it s.gaudino@dstech.it
Mikel Uriarte Óscar López Saioa Ros	P20 S21SEC	muriarte@s21sec.com olopez@s21sec.com sros@s21sec.com
Karolina Lach Bartosz Lipski	P22 CF	klach@cloudferro.com blipski@cloudferro.com
Marcin Oksimovic	P23 ELECTRUM	moksimowicz@electrum.pl
Walter Quadrini	P24 POLIMI	walter.quadrini@polimi.it
Dario LaCarrubba Maria Rossetti	P25 MADE	dario.lacarrubba@made-cc.eu maria.rossetti@made-cc.eu
Pablo Patús	P26 NASERTIC	ppatusdi@nasertic.es
Charly Gerber	P27 SIPBB	charly.gerber@sipbb.ch

History

Date	Version	Change
10-OCT-2022	0.1	Table of Contents and assignments
31-OCT-2022	0.2	Consolidated version after 1 st round of contributions
24-NOV-2022	0.3	Consolidated version after 2 nd and 3 rd round of contributions

2-DEC-2022	1.0	Final version submitted to the EC
------------	-----	-----------------------------------

Key Data

Keywords	SotA, state of the art, market analysis, continuum
Lead Editor	P19 – DS TECH
Internal Reviewer(s)	PIC

Table of contents

Table of contents	6
List of tables	8
List of figures	8
List of acronyms	11
1. About this document.....	18
1.1. Deliverable context	18
1.2. The rationale behind the structure.....	18
2. Introduction to aerOS	20
3. State of the art.....	22
3.1. Edge-cloud continuum orchestration	22
3.1.1. Smart networking and infrastructure management	22
3.1.2. Resource orchestration approaches.....	33
3.1.3. APIs, monitoring and communication services for the continuum.....	37
3.1.4. Data orchestration approaches	40
3.2. Review of relevant techniques for the meta operating system.....	43
3.2.1. Real-time containers in the Industry	44
3.2.2. Edge-native approaches: cloud-native techniques applied along the computing continuum...	45
3.2.3. Self-* capabilities of heterogeneous nodes.....	54
3.2.4. Data syntactic and semantic interoperability in the continuum	65
3.2.5. Data sovereignty, governance and lineage policies	68
3.2.6. Advanced AI management approaches.....	69
3.2.7. Security, integrity, trust, privacy and policy enforcement in the computing continuum	75
3.2.8. From DevOps to DevSecOps to DevPrivSecOps	78
3.2.9. Distributed multiplane analytics	82
3.3. Surrounding ecosystem	84
3.3.1. Industrial approach to edge-cloud continuum in Industry (I4.0 and I5.0)	84
3.3.2. Current existing standards related to aerOS.....	101
3.3.3. Review of the DATA-01-05 cluster.....	106
3.3.4. Other related projects.....	110
3.4. Review of current approaches in selected verticals	120
3.4.1. Edge-cloud technologies in robotics and manufacturing sector	120
3.4.2. Edge-cloud technologies in maritime port sector	123
3.4.3. Edge-cloud technologies in mobile machinery sector	126
3.4.4. Edge-cloud technologies in telecom operators sector (a usability perspective)	128
3.4.5. Edge-cloud technologies in containerised data centres close to renewable energy sources ..	134
4. Market analysis report	137
4.1. aerOS market.....	137
4.1.1. Target Market	137
4.1.2. Correlative Market.....	139
4.1.3. Market Size and growth.....	144
4.1.4. Market Trends.....	145
4.2. Influencers market factors.....	148

4.2.1.	Political	149
4.2.2.	Economical	149
4.2.3.	Social	150
4.2.4.	Technological.....	152
4.2.5.	Legal	152
4.2.6.	Environmental.....	154
4.3.	Competitive Landscape.....	155
4.3.1.	Relevant similar projects	155
4.3.2.	Business solutions.....	156
4.4.	aerOS market position.....	159
4.5.	Verticals addressed in aerOS - Market trends.....	161
4.5.1.	Manufacturing – production	161
4.5.2.	Renewable energy sources.....	165
4.5.3.	Port Continuum.....	168
4.5.4.	Smart Building.....	171
4.5.5.	Machinery of agriculture, forestry and construction.....	173
4.6.	Partners and Stakeholders engagement activities.....	177
4.6.1.	Interviews	178
4.6.2.	Focus Groups	178
4.6.3.	Written Interviews	178
4.6.4.	Online survey.....	178
4.6.5.	Workshop.....	188
5.	Conclusions	192
	References	193
A.	Interviews	216
B.	Focus groups	222
C.	Written interviews.....	231

List of tables

Table 1. Deliverable context.....	18
Table 2. Explainable AI tools	72
Table 3. Security, Privacy, Trust Challenges and Solutions.....	77
Table 4. Libraries and Tools for Distributed AI.....	83

List of figures

Figure 1. Structure of the scientific state of the art section.	19
Figure 2. Structure of the computing continuum – target of aerOS	20
Figure 3. Summary of aerOS objectives and approach	21
Figure 4. The concept of Network Virtualisation (source: [ITU3011])	23
Figure 5. Simplified view of the Cloud Network model applied to a hybrid satellite/terrestrial network infrastructure	23
Figure 6. RESTful APIs for the Service Based Interfaces and Northbound communication.....	27
Figure 7. ETSI NFV architecture	29
Figure 8. Network functions evolution [SNIM-8].....	31
Figure 9. Third-party Standalone (A) and Non-standalone (B) NetApp representation	33
Figure 10. NetApp's interaction with the data and control plane when a Vertical application is provided.....	33
Figure 11. NFV Reference Architectural Framework [ROA-3].....	34
Figure 12. Graph representation of an end-to-end network service [ROA-3]	34
Figure 13. Resource management and orchestration across multi-technological and administrative domains.	35
Figure 14. An architecture for multi-domain resource orchestration for network slicing [ROA-8].....	35
Figure 14. Metropolis model structure [SO-16]	38
Figure 16. ROS contributors [SO-21].....	39
Figure 16. ROS Ecosystem PR governance [SO-18]	39
Figure 15. Architecture of the data fabric.....	41
Figure 16. Simplified operating model of data mesh [DOA-12].....	43
Figure 17. BalenaOS block architecture [ENA-4].....	46
Figure 18. Evolution of embedded systems from the point of view of Pantavisor [ENA-5]	47
Figure 19. Architecture comparison between traditional container-based and Pantavisor-based [ENA-6]	47
Figure 20. KubeEdge architecture [ENA-11]	49
Figure 21. Main Open Horizon components [ENA-14]	50
Figure 22. Baetyl architecture [ENA-15]	50
Figure 23. Baetyl architecture [ENA-15]	51
Figure 24. Google distributed cloud architecture [ENA-25]	52
Figure 25. Differences between Kata Containers' VMs and traditional containers [ENA-26]	53
Figure 26. Creating and running a Unikernel [ENA-28].....	53
Figure 27. Architecture of the Docker Engine for running Wasm workloads [ENA-35]	54
Figure 28. Phases of the feedback loop "MAPE-K" [SELF-24].	57
Figure 29. Examples of vehicle detection in the traffic surveillance system [SELF-39].	59
Figure 30. Communication network for real-time ship monitoring [SELF-44].	60
Figure 31. Example of system grid partition [SELF-56].....	61
Figure 32. 5G network model with hexagonal cells, divided into three sectors [SELF-62]	62
Figure 33. Self-adaptation model for microservice systems [SELF-71]	63
Figure 34. Networks of the self-adaptive system [75].....	64
Figure 35. Propulsion system of an electric vehicle [SELF-78].....	64
Figure 36. Remote self-learning driving cycle [SELF-81].....	65
Figure 37. Main aspect of data governance [DSGP-1].....	68
Figure 38. Evolution from DevOps to DevSecOps to DevPrivSecOps.....	78
Figure 39. DevOps workflow	79
Figure 40. security Controls in the DevSecOps workflow	80

Figure 41. Privacy considerations in the DevPrivSecOps workflow	81
Figure 42. Industry 3.0 most important technological advances.	85
Figure 43. Industry 4.0 most important technological advances.	86
Figure 44. Technology Enablers of I5.0 [IECC-1]	86
Figure 45. Industry 5.0 compared to Industry 4.0 [IECC-1]	87
Figure 46. Reference Architecture Model for Industry 4.0 [IECC-3]	88
Figure 47. Product Life Cycle axis.	89
Figure 48. Architecture layers.	89
Figure 49. Functional viewpoint of IIRA (a) and RAMI4.0 (b)	90
Figure 50. The open Industry 4.0 Alliance Technical Architecture [IECC-9].	91
Figure 51. The open edge computing layer.	92
Figure 52. Open operator cloud.	92
Figure 53. The common cloud central.	93
Figure 54. IDSA-RAM general structure [IECC-8].	95
Figure 55. Representation of the information model of IDSA-RAM [IECC-8].	96
Figure 56. FIWARE Smart Industry Reference Architecture [IECC-4]	97
Figure 57. Digital Factory Alliance Reference Architecture for Industry 4.0 [IECC-7]	99
Figure 58. NGSi-LD meta-model [CES-6]	103
Figure 59. NGSi-LD distributed architecture	103
Figure 60. Repositories and domains in Smart data models	104
Figure 61: Meta-Operating Systems for the next generation IoT and Edge Computing - Source: Factsheet for Horizon Europe, Cluster 4, Destination 3: "Future European Platforms for the Edge: Meta-Operating Systems"	106
Figure 62. Overview of the 5G-PPP Programme	116
Figure 63 Mapping of use cases to vertical categories	116
Figure 64. 5G Infrastructure PPP Phase 3 Platforms Projects – Geographic Cartography	117
Figure 65. Vertical industries under validation by ICT-17 and ICT-19 projects	118
Figure 66. 5G-PPP Phase 3 Reference Figure	119
Figure 67. 5G-PPP Key Achievements v3.0	119
Figure 68. The automation pyramid	120
Figure 69. "Tilted" pyramid	122
Figure 70. Dell-Intel edge and IoT portfolio for port operations [EMP-6].	124
Figure 71: Overview of the Involvement of SDOs for Edge Computing in Mobile Networks, inspired by [ETS-9]	130
Figure 72: Simplified View of the 3GPP WGs Edge Work	131
Figure 73: Synergised Mobile Edge Cloud Architecture Supported by 3GPP and ETSI ISG MEC specifications [ETS-9]	132
Figure 74: direct/physical power purchase agreement (PPA) [ERE-3]]	135
Figure 75: Use of cloud computing services, 2020 and 202. Source: Eurostat, Cloud computing - statistics on the use by enterprises, 2021	138
Figure 76. Edge Computing Submarkets. Source: Gartner, Market Guide for Edge Computing, 2022	139
Figure 77: Estimated revenues in EU 2025 (€B). Source 1: European Telecommunications Network Operators' Association (Etno), Connectivity & Beyond: How Telcos Can Accelerate a Digital Future for All, 2021	143
Figure 78 - Global Blockchain Market Share, by industry, 2021. Source: Fortune business insights, Blockchain market size, share & Covid-19 impact analysis, 2022	144
Figure 79. Edge computing potential value by 2025. Source: McKinsey & Company (JM Chabas, C. Gnanasambandam, S. Gupte, and M. Mahdavian), New demand, new markets: what edge computing means for hardware companies, 2018	145
Figure 80: Cloud computing in 2027. Source: Own elaboration based on Gartner, The Future of Cloud Computing in 2027: From Technology to Business Innovation, 2022	146
Figure 81. Phases of Edge Computing Market Focus. Source: Gartner, Market Guide for Edge Computing, 2022	147
Figure 82: Services Comparison between aerOS and the HE projects ICOS, FLUIDOS and NEMO. Source: Own elaboration	156
Figure 83: Pilot Comparison between aerOS and the HE projects ICO. Source: Own elaboration	156

Figure 84. Surveyed manufacturers plan to focus on a range of technologies to increase operational efficiencies over next 12 months. Source: Deloitte, 2023 Deloitte manufacturing outlook survey, 2022	163
Figure 85. Estimated economic value by use case, 2020–30, \$ billions. Source: McKinsey, The Internet of Things: Catching up to an accelerating opportunity, 2021	164
Figure 86:	166
Figure 87. Share of energy from renewable sources, 2020 (% of gross final energy consumption) Source: Eurostat, Renewable energy statistics, 2022	166
Figure 88. The forecast of worldwide electricity consumption, for the period of 2020 to 2050. Source: Minh, Q.N.; Nguyen, V.-H.; Quy, V.K.; Ngoc, L.A.; Chehri, A.; Jeon, G. Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. Energies 2022	167
Figure 89: Worldwide container port traffic (2000-2020) – TEU (Twenty-foot Equivalent Unit). Source: The World Bank, Container port traffic.....	168
Figure 90: The layout of an automated container terminal. Source: Yang, Yongsheng & Zhong, Meisu & Yao, Haiqing & YU, Fang & Fu, Xiuwen & Postolache, Octavian. Internet of things for smart ports: Technologies and challenges. IEEE Instrumentation & Measurement Magazine, 2018	169
Figure 91: A comprehensive building optimisation ecosystem. Source: Deloitte, smart buildings – four considerations for creating people-centred smart, digital workspaces, 2018	172
Figure 92: Smart forestry application. Source: Business Finland – Mediabank, Finnish solutions for smart forestry.....	175
Figure 93: Estimated economic value by use case for construction, 2020–30, \$ billions. Source: McKinsey and Company, The Internet of Things: Catching up to an accelerating opportunity, 2021	176
Figure 94: Estimated economic value by use case for farms, 2020–30, \$ billions. Source: McKinsey and Company, The Internet of Things: Catching up to an accelerating opportunity, 2021	177
Figure 95: aerOS survey participants general characteristics. Source: Own elaboration	181
Figure 96: Business areas that could benefit the most from aerOS and related solutions. Source: Own elaboration	182
Figure 97: Awareness of EtC solutions on the market; Position towards Service Brochure on Cloud and Edge Provisioning. Source: Own elaboration.....	182
Figure 98: Level of diffusion of AI and blockchain technologies, and the latter estimated improvements in data and data providers certification. Source: Own elaboration	183
Figure 99: Major concerns and challenges regarding IoT and Edge solutions deployment. Source: Own elaboration.....	183
Figure 100: Perceived concerns of political issues related to the adoption of Cloud and Edge systems, and of a European vs Non-European Cloud by EU companies. Source: Own elaboration.....	184
Figure 101: Influence of social factors like the availability of a skilled workforce on the adoption of Edge Computing technologies. Source: Own elaboration	184
Figure 102: Data relevance in business and operational decisions; data scientist main activities; most required features for data storage systems. Source: Own elaboration	185
Figure 103: Technical aspects related to the presence of heterogeneous sources of information. Source: Own elaboration.....	185
Figure 104: Perception of Zero Latency necessity; SSOT company reference; successful prediction model influence on correlative geographical location. Source: Own elaboration.....	186
Figure 105: Privacy concerns for data sharing; bottlenecks related to huge amount of data; perceived security in letting data travel to cloud-based nodes outside the network. Source: Own elaboration	187
Figure 106: System Evaluation regarding intelligence from cloud to on premises; evaluation of company deployed applications and services based on their staff user friendliness; main devices that implement the edge infrastructure of the participant. Source: Own elaboration	187
Figure 107: Evidences of the conclusions extracted out of the D2.1	192

List of acronyms

Acronym	Explanation
5G EVE	European 5G validation platform for extensive trials
5G NEC	5G Network Exposure Function
5G NetApp	5G Network App
5GAP	5G Action Plan for Europe
5GC	5G Core
5G-PPP	5G Infrastructure Public Private Partnership
ABAC	Attribute-Based Access Control
ADAS	Advanced Driver Assistance System
AF	Application Function
AGV	Automated Guided Vehicles
AI	Artificial Intelligence
AL	Active Learning
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
AR	Augmented Reality
AV	Audio-visual
AWS	Amazon Web Services
B2B	Business to business
B5G	Beyond 5G
BGP	Border Gateway Protocol
BSS	Business Support Systems
CAGR	Compound Annual Growth Rate
CAM	Connected and Automated Mobility
CAN	Controller Area Network
CAPEX	Capital expenditure or capital expense
CAPIF	Common API Framework
CBS	Constant-bandwidth server
CD	Continuous Delivery
CEO	Chief Executive Officer
CI	Continuous Integration
CLA	Closed-loop automation
CNCF	Cloud Native Computing Foundation
CNF	Cloud-native Network Function
CoT	Cloud of Things
COTS	Commercial off-the-shelf
CPD	Consumer Purchase Data
CPPS	Ciber Physical Production Systems
CPS	Cyber Physical System
CPU	Central Processing Unit
CSA	Coordination and support actions
CSP	Communication Service Provider

D2D	Device to device
DaaP	Data-as-a-Product
DC	Data Centre
DCPS	Data-Centric Publish-Subscribe
DDoS	Distributed Denial of Service
DDS	Data Distribution Service
DetNet	Deterministic Networking
DevOps	Development and Operations
DevSecOps	Development, Security and Operations
DevSecPrivOps	Development, Security, Privacy and Operations
DFA	Digital Factory Alliance
DGA	Data Governance Act
DLT	Distributed Ledger Technology
DMP	Digital Manufacturing Platform
DNS	Domain Name System
DRL	Deep Reinforcement Learning
DSS	Decision Support System
E2E	End to end
EC	European Commission
EC	Edge-Cloud
ECaaS	Edge-Compute-as-a-Service
ECC	Edge Computing Consortium
EDF	Earliest Deadline First
EDGEAPP	Architecture for Enabling Edge Applications
EDN	Edge Data Network
EE	Energy Efficiency
EHS	Employee Health and Safety
EMF	Electromagnetic Field
ERD	Entity Relationship Diagram
ESB	Enterprise Service Bus
ESG	Environmental, social and governance
ETSI	European Telecommunications Standards Institute
ETSI ENI ISG	ETSI Experiential Network Intelligence ISG
ETSI ZSM ISG	ETSI Zero Touch network and Service Management ISG
EU	European Union
EVE	Edge Virtualization Engine
FaaS	Function-as-a-Service
FiWi	FiberWireless
FoF	Factory of the Future
FRMCS	Future Railway Mobile Communication System
FWA	Fixed Wireless Access
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit

GSMA	Global System for Mobile Communications Association
H2020	Horizon 2020 EU's research and innovation funding programme from 2014-2020
HA	High Availability
HBA	Home and Building Automation
HCP	Handover control parameter
HCRM	Home Cognitive Resource Manager
HD	High Definition
HPC	High Performance Computing
HPCP	High Performance Computing Platform
I3.0, I4.0, I5.0	Industry 3.0, 4.0, 5.0
IaaS	Infrastructure-as-a-Service
ICM	Interactive cell model
ICOS	IoT to Cloud Operating System
ICT	Information and Communication Technology
IdM	Identity Management
IDS	Intrusion Detection System
IDS	Industrial Data Space
IDSA	International Data Spaces Association
IE	Infrastructure Element
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IIoT	Industrial Internet of Things
IIRA	Industrial Internet Reference Architecture
ILP	Integer Linear Programming
IoT	Internet of Things
IP	Internet Protocol
iPaaS	Integration-Platforms-as-a-Service
IPR	Intellectual Property Rights
ISG	Industry Specification Group
ISO	International Organization for Standardization
ISP	Internet Service Provider
IT	Information Technology
ITU-T	Telecommunication Standardization Sector of the International Telecommunication Union
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
K8s	Kubernetes
KDT	Key Digital Technology
KPI	Key Performance Indicator
LAN	Local Area Network
LCIM	Levels of Conceptual Interoperability Model classification
LD	Linked Data
LF	Linux Foundation
LIDAR	Light Detection and Ranging

LINP	Logically Isolated Network Partitions
LPG	Labelled Property Graph
LXC	LinuX Containers
MAC	Mandatory Access Control
MANO	NFV management and orchestration
MAPE	Monitor, Analyse, Plan, Execute
MEC	Multi-access Edge Computing
MEP	Multi-access Edge Computing Platform
MES	Manufacturing Execution System
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MMTC	Massive Machine Type Communication
MNO	Mobile Network Operator
ModelOps	Model Operations
MOM	Message-Oriented Middleware
MPLS	Multiprotocol Label Switching
MQTT	MQ Telemetry Transport
MRS	Modular Robotic System
MWC	Mobile World Congress
NaaS	Network-as-a-Service
NAO	NetApp Orchestration
NB-IoT	Narrow Band Internet of Things
NEF	Network Exposure Function
NEMO	Next Generation Meta Operating system
NETCONF	Network Configuration Protocol
NFV	Network Functions Virtualisation
NFVI	NFV Infrastructure
NFVO	NFV Orchestrator
NGSI	New Generation Service Interface
NI	Network Intelligence
NIST	National Institute of Standards and Technology (USA)
NLP	Natural Language Processing
NoSQL	Non-SQL or not relational
NPN	Non-Public Network
NRF	Network Repository Function
NS	Network Service
NSaaS	Network-Slice-as-a-Service
NVE	Network Virtualization Edge
NWDAF	Network Data Analytics Function
OCI	Open Container Initiative
OEM	Original Equipment Manufacturer
OH	Open Horizon
OI4.0	Open Industry 4.0 Alliance
OLTP	OnLine Transaction Processing

OODA	Observe, Orient, Decide, Act
OPAG	Operator Platform API Group
OPC	Open Platform Communication
OPEX	Operating expenditure or operating expense
OPG	Operator Platform Group
O-RAN	Open RAN
OS	Operating System
OSS	Operations Support Systems
OSS	Open-Source Software
OT	Operational Technology
OWL	Web Ontology Language
P2P	Peer-to-peer
PaaS	Platform-as-a-Service
PAP	Personal Auto Policy
PdM	Predictive Maintenance
PDP	Policy Decision Point
PEP	Performance-Enhancing Proxy
PIA	Privacy Impact Assessment
PII	Personally Identifiable Information
PLC	Programmable Logic Controller
PMSE	Programme Making and Special Events
PNF	Physical Network Function
PoC	Proof of Concept
PPA	Power Purchase Agreement
PPDR	Public Protection and Disaster Relief
PSO	Particle Swarm Optimization
QoS	Quality of Service
R&D	Research and Development
R&I	Research and Innovation
RA	Reference Architecture
RAM	Random Access Memory
RAMI 4.0	Reference Architecture Model for Industry 4.0
RAN	Radio Access Network
RBAC	Role-Based Access Control
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
REST	Representational State Transfer
RIA	Research and Innovation Action
RIS	Reconfigurable Intelligent Surface
ROS	Robot Operating System
RT	Real-Time
RTAI	Real-Time Application Interface
RTU	Remote Terminal Unit
SA	Stand-alone

SaaS	Software-as-a-Service
SAE	Society of Automotive Engineers
S-AIS	Satellite Automatic Identification System
SBA	Service Based Architecture
SBI	Service Based Interface
SCM	Self-sufficient cell model
SDG	Sustainable Development Goal
SDN	Software Defined Networking
SDO	Standards Developing Organization
SD-SEC	Software-Defined Security
SD-WAN	Software-Defined Wide Area Network
SEAL	Service Enabler Architecture Layer for Verticals
SLA	Service Level Agreements
SLAM	Simultaneous Localisation And Mapping
SME	Subject Matter Expert
SNS	Smart Networks Service
SOAP	Simple Object Access Protocol
SotA	State of the Art
SQL	Structured Query Language
SSL	Secure Sockets Layer
SSL	Semi-supervised learning
T&L	Transport and Logistics
TEC	Telco Edge Cloud
TEE	Trusted Execution Environment
TLS	Transport Layer Security
TMS	Terminal Management System
TRiSM	Trust, risk and security management
TSN	Time Sensitive Networking
TSP	Technology and Service Provider
TTEthernet	Time-Triggered Ethernet
UDP	User Datagram Protocol
UHD	Ultra-high Definition
UML	Unified Modelling Language
UN	United Nations
V2G	Vehicle-to-grid
V2I	Vehicle to Infrastructure
VAE	Vertical Application Enabler
VIM	Virtualized Infrastructure Manager
VLAN	Virtual Local Area Networks
VM	Virtual Machine
VNF	Virtualised Network Function
VNFM	VNF Manager
VNI	VXLAN Network Identifier
VoD	Video on Demand

VPN	Virtual Private Network
vRAN	Virtual Radio Access Network
VXLAN	Virtual Extensible LAN
W3C	World Wide Web Consortium
WAN	Wide Area Network
WASI	Web Assembly System Interface
Wasm	WebAssembly
WoT	Web of Things
WSN	Wireless Sensor Network
WWW	World Wide Web
XACML	eXtensible Access Control Markup Language
XAI	Explainable AI
XML	Extensible Markup Language
XSD	XML Schema Definition
YANG	Yet Another Next Generation
ZDM	Zero-Defect Manufacturing
ZTM	Zero-Touch Management

1. About this document

The main objective of this document is **to realise the current status of techniques, technologies and methodologies** related to forthcoming aerOS innovations **and to analyse the status of the market** (both the niche of the meta operating system for the continuum in general and also focusing specifically on the segments of aerOS' pilots). Considering that the larger bulk of innovations, as well as the workload to be exerted in the project is focused on the technologies surrounding the concept of meta operating system for the continuum, this deliverable is considered paramount to establish a solid baseline to advance beyond. D2.1 will be used during the rest of the project as the background of the status of different enabling technologies, allowing partners to select the best alternatives to develop the results over. In addition, it allows the project to position itself within the research landscape since very early stages of execution.

1.1. Deliverable context

Table 1. Deliverable context

Item	Description												
Objectives	This deliverable is directly related with all objectives of aerOS but O6 and O7. Objectives O1 to O5 are related to different techniques, technologies or methodologies that aerOS is progressing beyond the state of the art (optimal orchestration, smart network functions, decentralised security, privacy and trust, distributed explainable AI components and data autonomy strategy correspondingly). This deliverable (D2.1) conforms the keystone around which those innovations will orbit, as it has the goal to inspect the current trends and advances on those specific fields. Acknowledging the status of the research and deployment will allow for a more efficient work towards improving them.												
Work plan	Deliverable D2.1 is the first technical deliverable of the project. It is the first deliverable of WP2, that will be followed by other documents such as the description of use cases or the architecture of the project. Deliverable D2.1 is completed at the end of the third month of the project, where only three WPs had started (WP1, WP2 and WP6). It serves as the milestone to kick-off technical activities of the project in WP3 and WP4. At the same time, it is also related to the pilots (WP5) in terms of analysing the current technology and status of the market of the different verticals of the project.												
Milestones	<p>The submission of deliverable D2.1 is directly related to the completion of milestone MS1: Identity definition. With D2.1, MS1 is fully achieved, as D1.1 was already submitted by M1 and the website of the project has been active and online since M1 as well.</p> <table><tr><th>Milestone No</th><th>Milestone Name</th><th>Work Package No</th><th>Lead Beneficiary</th><th>Means of Verification</th><th>Due Date (month)</th></tr><tr><td>1</td><td>Identity definition</td><td>WP6, WP1</td><td>15-INFOLYSIS</td><td>Web site, State of the art and Kick-off meeting (D1.1, D2.1)</td><td>3</td></tr></table>	Milestone No	Milestone Name	Work Package No	Lead Beneficiary	Means of Verification	Due Date (month)	1	Identity definition	WP6, WP1	15-INFOLYSIS	Web site, State of the art and Kick-off meeting (D1.1, D2.1)	3
Milestone No	Milestone Name	Work Package No	Lead Beneficiary	Means of Verification	Due Date (month)								
1	Identity definition	WP6, WP1	15-INFOLYSIS	Web site, State of the art and Kick-off meeting (D1.1, D2.1)	3								
Deliverables	D2.1 is not directly fed from any other previous deliverables. It is expected to serve as a baselines for the forthcoming technical deliverables: D2.3, D2.5, D2.6, D2.7, D3.1, D4.1 and D5.1.												
Risks	#4 -Change of project requirements due to evolution of relevant technology and market landscape: D2.1 establishes a baseline to ensure that technical activities will start from the most recent analysis of existing technologies and trends.												

1.2. The rationale behind the structure

The content of the deliverable is organized in two main sections (alongside a short introduction and a conclusion section), aligned with the scope of the task T2.1.

- **Section 2.** This section introduces the reader to aerOS project in a light fashion, exposing its objectives and main proposals.
- **Section 3:** This is one of the core sections of the document. It reports the findings by aerOS partners about the technologies and techniques related to aerOS scientific scope. This session is subsequently divided in four sub-sections. The first one is devoted to the main aspects of orchestration required in the IoT-edge-cloud continuum, that have been catalogued in four: network, resources, services and data. The second sub-section involves a myriad of technological flavours. It is structured in 9 sub-chapters, that map (almost directly) the target goals of the tasks in WP3 of WP4. Third, a state of the art of a research project is not complete without analysing other research actions that address the same or very similar topics; in this subsection, an overview is made of the projects in the cluster of aerOS (that coming of DATA-01-05-2021 call) and other initiatives like 5GPP or Industrial architectures promoted by relevant clusters. In addition, this third sub-section also gathers the most relevant standardisation actions that work on related aerOS technologies as of today. Last, the fourth sub-section deals with the status of the technology in the different pilots of aerOS (5).

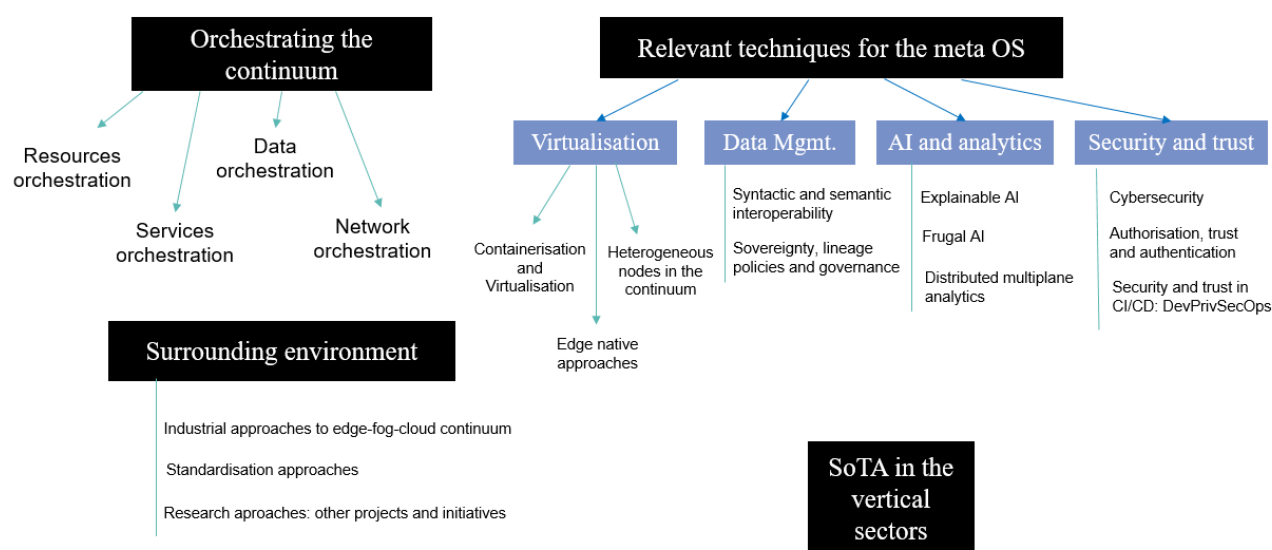


Figure 1. Structure of the scientific state of the art section.

- **Section 4:** This is the other core section of the document. It gathers all the advances and data obtained after analysing the market of aerOS. In this section, the market of both the meta OS and the pilots of the project are overviewed. It is ulteriorly divided in 6 sub-sections. First of them focuses on a pure market analyses in terms of target, size, growth, brief overview of main competitors and trends. The second sub-sections drifts from the first as it devises a PESTEL analysis of the surrounding business environment to aerOS. It analysis the market from the political, economical, social, technological, environmental and legal aspects. Afterwards, the third sub-section digs deeper into the competitive analysis of aerOS, both from existing products perspective and from business solutions on the different core technical domains covered by aerOS (cloud computing, edge computing AI and IoT). Fourth section compiles the findings of the three previous chapters and makes the first attempt of exactly positioning aerOS in the market. The fifth sub-section provides a thorough review on the status of commercial and open source solutions and the associated companies (the market) of the specific segments of aerOS pilots (manufacturing, renewable energy sources, port continuum, smart building and machinery of agriculture, forestry and production). The last sub-section condenses the results and conclusions obtained conducting several activities conducted to gather more insights on the market of aerOS. In particular, it includes the results of interviews with internal and external experts, the focus groups with specific set of entities, the written interviews that were submitted to judiciously selected respondents and the statistics of the online survey published as well as the comments and global conclusions of the online workshop held online on November 29th, 2022.
- **Section 5:** The document concludes with a conclusion section, followed by the list of consulted references.

2. Introduction to aerOS

The unprecedented data explosion and the evolving capabilities of virtual infrastructures, set the scene for developing a new paradigm for data and compute resource management in EU. Rapidly increasing data volumes necessitate application developers and service providers to leverage data processing capabilities offered by segmented compute infrastructures, including all edge tiers (far, micro, etc.) up to the cloud. Processing needs to be performed closer to the data sources (often smart devices), in an effort to minimise latency, save bandwidth, improve security, guarantee privacy and increase autonomy. However, this requires highly efficient and real-time responding distributed edge; an especially challenging task, due to the heterogeneity and diversity of involved technologies and because of existing legacy investments. To achieve scalable and long-term evolving solution(s), high complexity of distributed edge has to be managed by supporting variety of (current and future) deployment models and open standards.

aerOS overarching goal is to design and build a virtualised, platform-agnostic **meta operating system for the IoT edge-cloud continuum**. As a solution, to be executed on any Infrastructure Element within the IoT edge-cloud continuum – hence, independent from underlying hardware and operating system(s) – **aerOS** will: (i) deliver common virtualised services to enable orchestration, virtual communication (network-related programmable functions), and efficient support for frugal, explainable AI and creation of distributed data-driven applications; (ii) expose an API to be available anywhere and anytime (location-time independent), flexible, resilient and platform-agnostic; and (iii) include a set of infrastructural services and features addressing cybersecurity, trustworthiness and manageability. **aerOS** will: (a) use context-awareness to distribute software task (application) execution requests; (b) support intelligence as close to the events as possible; (c) support execution of services using “abstract resources” (e.g., virtual machines, containers) connected through a smart network infrastructure; (d) allocate and orchestrate abstract resources, responsible for executing service chain(s) and (e) support for scalable data autonomy.

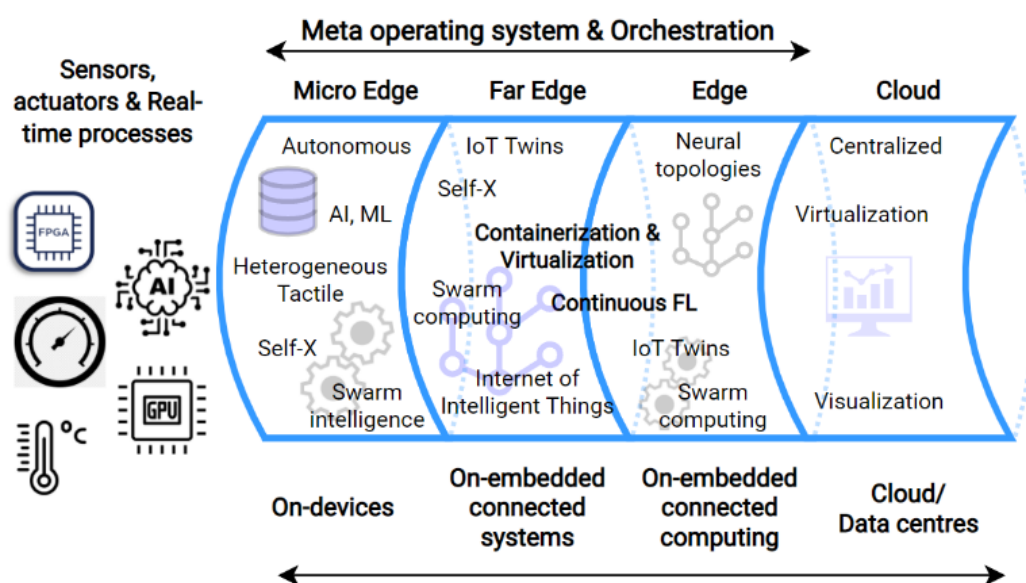


Figure 2. Structure of the computing continuum – target of aerOS

Moreover, **aerOS** will leverage **European leadership** in automation systems in industry (where edge resides) and pointedly prove how **European industry** can benefit from decentralised, platform-agnostic IoT edge-cloud continuum data-processing ecosystem, to build competitive advantages e.g., reduced time to decisions; cost and time efficient, secure, trustworthy data sharing and control; semi-autonomous action taking; agile operations; sustainable, human-centric data processing, governance, and interoperability; reduced external traffic; and improved latency. The **aerOS** approach will be generic and directly **applicable to any vertical, cross-vertical** business process, and several different physical or virtual platforms. It will answer the urgent need for a trustworthy, decentralised, autonomous, orchestrated solution, enabling bottom-up formation of compute continuum ecosystems, where hyper-distributed applications will be efficiently executed, within any selected “fragment” of heterogeneous physical infrastructure

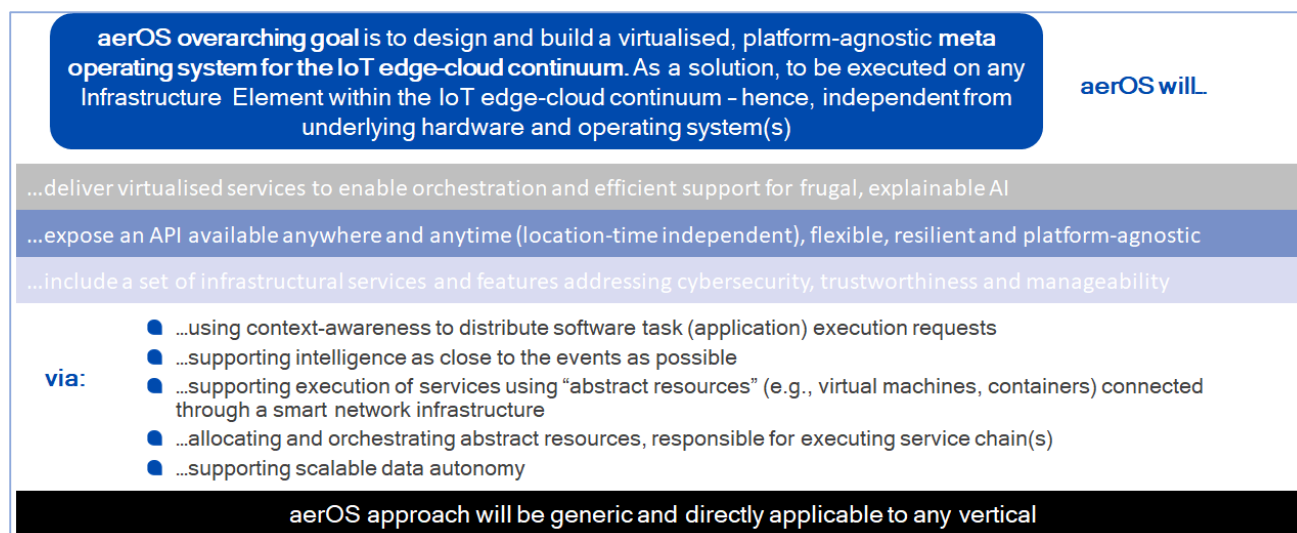


Figure 3. Summary of aerOS objectives and approach

The figure above summarises the global approach of aerOS as a Research and Innovation action covering the following goals:

- O1: Design, implementation and validation of aerOS for optimal orchestration
- O2: Intelligent realisation of smart network functions for aerOS
- O3: Definition and implementation of decentralised security, privacy and trust
- O4: Definition and implementation of distributed AI components with explainability
- O5: Specification and implementation of a Data Autonomy strategy for the IoT edge-cloud continuum
- O6: Definition, deployment, and evaluation of real-life use cases (5)
- O7: Global ecosystem creation, maximisation of impact and Open Call conduction

In order to achieve those goals, aerOS aims at evolving capabilities of the computing fabric across the IoT-edge-cloud continuum, specially paying attention to the following aspects:

- Orchestration of resources, services, data and network in the continuum
- Data management, sovereignty, governance and lineage policies.
- Smart Networking:
- Self-* capabilities of heterogeneous node in the continuum
- Pub/Sub Broker Support
- Frugal AI with Explainability (FAI):
- TSN Support:
- Benchmarking Tools
- Resilience Policies and Mechanisms
- Embedded Analytics
- Cybersecurity Tools
- Privacy
- Trust
- Management frameworks for the continuum

Finally, both the exposed objectives and the evolution of the mentioned technologies will be validated by deploying a series of real-life use-cases divided in five different pilots:

- Data-Driven Cognitive Production Lines (Manufacturing Autonomy Level 4 – MAL4)

- Containerised Edge Computing near Renewable Energy Sources
- High Performance Computing Platform for Connected and Cooperative Agricultural Mobile Machinery to Enable CO2 Neutral Farming (HPCP-F)
- Smart edge services for the Port Continuum
- Energy Efficient, Health Safe & Sustainable Smart Buildings

The goal of this document is to cover the current status of the above technological fields and application sectors so that the project will have a solid baseline to research upon and advance.

3. State of the art

The content included in this section is the result of an intensive research activity carried out by aerOS partners across different fields and aspects. It focuses on a scientific, comprehensive analysis of the current technologies and approaches that dominate each of the relevant domains of the project.

The structure responds to a judicious classification of the topics based on their technical proximity and to the relation from a workplan perspective, as it has been exposed in 1.2.

The methodology that has been followed to conduct this state of the art has been the following:

- Analysis of the different technical domains to be covered in aerOS and appointing of responsables according to expertise and workplan.
- Structuration of the list of technical topics into logical grouping.
- Investigation of the current status of technologies in the different fields.
- Discussion among the partners and experts.
- Establishing a common ground of understanding in terminology.

In the following subsections, the results of such exercise are evidenced. The depth and length of the subsections relates to the different degree of implication toward the final results of the project.

3.1. Edge-cloud continuum orchestration

The goal of this section is to offer a global overview of the relevant concepts of the main challenge of aerOS: orchestration of network, data, resources and services in the continuum. The following sub-sections cover each of those four different paths. In addition, some of those are subsequently divided in inner sub-chapters in order to ease readability and comprehension of the content.

3.1.1. Smart networking and infrastructure management

In the context of this study, the term “smart networking” refers to the virtualization and abstraction of network resources (i.e., links, nodes and functionalities) and their provision to the end-user as-a-Service, in a cloud-like manner, featuring dynamic resource pooling and elasticity. In this context, the physical infrastructure is divided into a number of independent, logically isolated virtual networks (referred to as “slices”) that are made available to clients and renters. Although network slices may transcend many heterogeneous network domains (wired and wireless), the tenant “sees” and operates a single end-to-end virtualized service, and is unaware of the specifics of the underlying infrastructure.

Figure 4 depicts this concept in accordance with ITU-T Rec. Y3011 which refers to these slices as “Logically Isolated Network Partitions” (LINP). As seen, physical resources are converted into virtual resources and then combined to create virtual networks (LINPs).

The ability to insert traffic processing services in the network slice in the form of software virtual network appliances (or -more commonly- Virtual Network Functions/VNFs) is another added-value feature of the cloud network model, which has been specifically highlighted over the last few years with the advent of Network Functions Virtualisation (NFV). In this situation, VNFs like virtual firewalls, caches, media processors, deep packet inspectors, etc. can further improve a network slice.

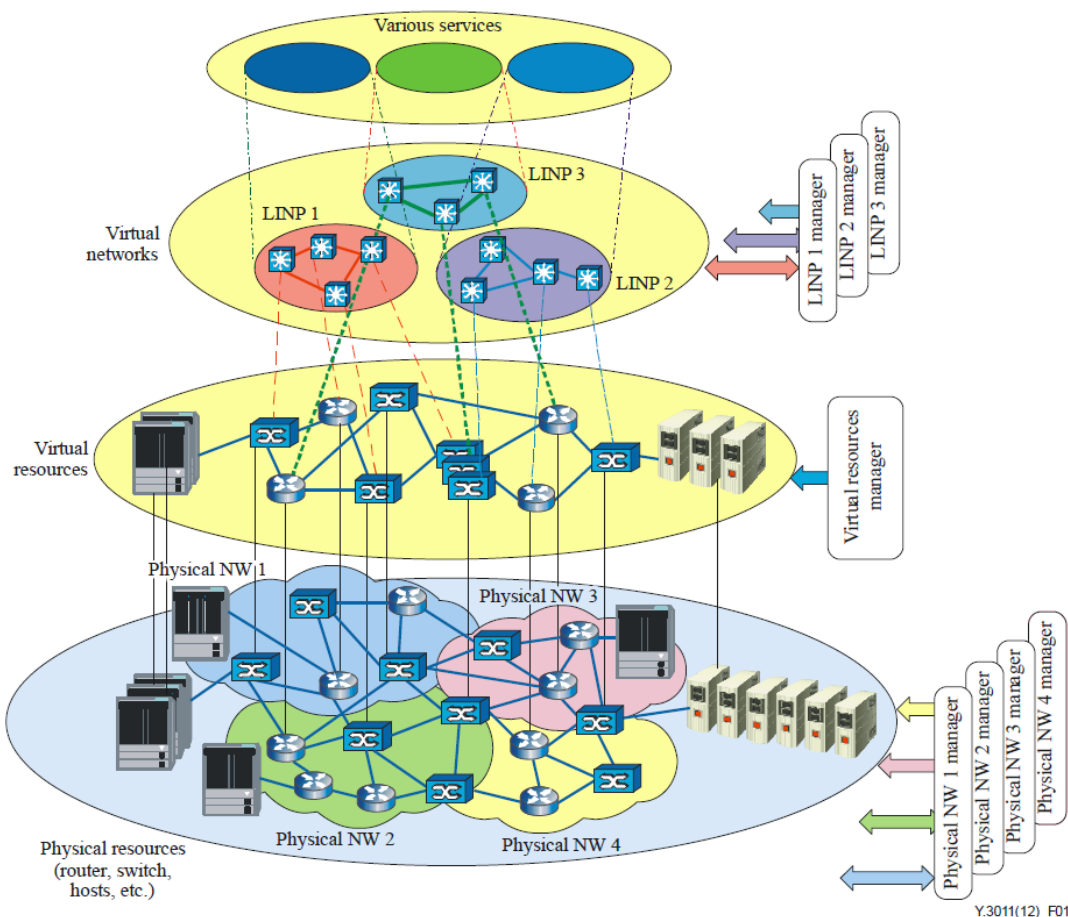


Figure 4. The concept of Network Virtualisation (source: [ITU3011])

In order to provide next-generation virtualized edge-to-cloud continuum, the cloud network model uses unique infrastructure management paradigms based on resource virtualization and federation across diverse physical infrastructures. These services consist of a connectivity element (virtual network slice), which may be upgraded with virtual network functions available on demand.

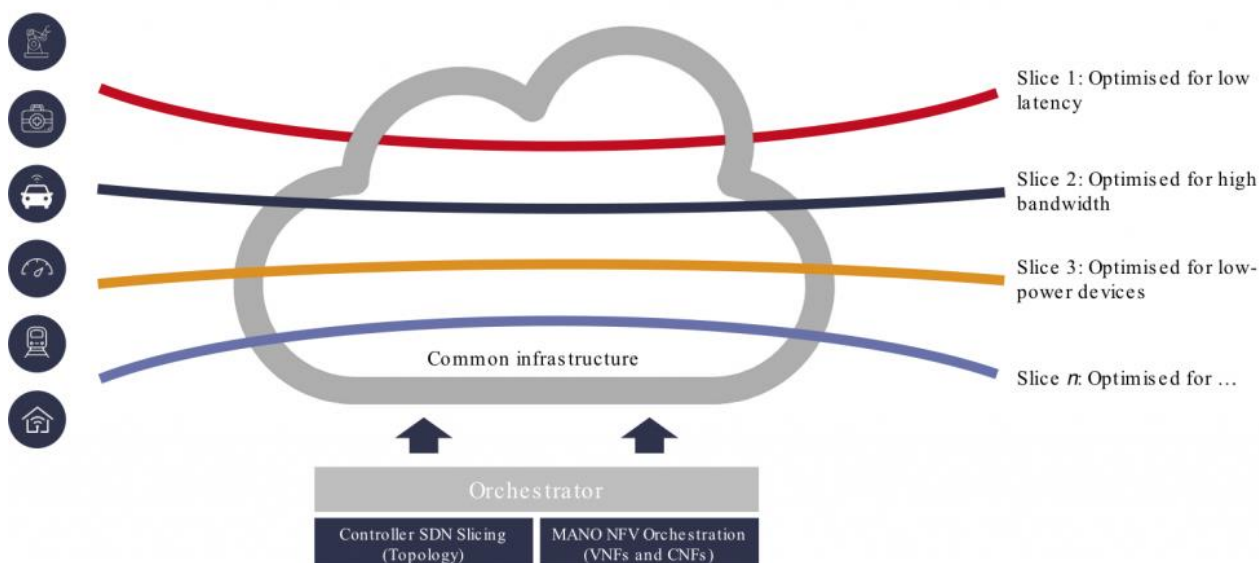


Figure 5. Simplified view of the Cloud Network model applied to a hybrid satellite/terrestrial network infrastructure

It should be mentioned that cloud networks' service offerings are significantly more comprehensive than those of current VPN bundles. Cloud Network Services offer full resource elasticity (i.e., up/down scaling) and can therefore support flexible Service Level Agreements (SLAs) and billing models based on usage thanks to the state-of-the-art technologies involved (such as network programmability and network functions virtualization to be discussed later in this deliverable). They can also natively enable connection and QoS in addition to a variety of rich in-network functionalities (VNFs), as was already described.

Modern computing Infrastructure-as-a-Service systems can be directly compared to this improved service delivering capacity. Users can request and purchase Virtual Machines (VMs) with pre-defined compute, memory, and storage capabilities through an IaaS service. Modern cloud computing systems also enable dynamic up- and down-scaling of VM resources in accordance with usage.

In a similar vein, users will be able to choose the virtual topology on a future satellite/terrestrial cloud network platform that best suits their needs in terms of endpoint/Point-of-Presence location, capacity, QoS, and in-network functionalities. These services will be provided as logically isolated services, transparently spanning the terrestrial and satellite domains. The specifics and architecture of the underlying physical infrastructure are concealed, and the user manages and keeps track of this cloud network service as if it were a standalone physical unified network.

3.1.1.1. Network infrastructure virtualization

Both in the internals of the data center (from now on, DC) and in the network that provides the interconnections for the DC, today the methods of network management that rely heavily on manual interventions are no longer appropriate. Higher levels of automation are required, along with flexibility to networking requirements.

In addition to the previously noted flexibility needed to adapt the network to actual conditions, the DC interconnections must scale to handle high tenant demand. This is a result of the growing virtualization technologies made available by the computing industry, which enable effective resource sharing.

The DC's multi-tenancy system enables users to share information resources. However, access to those resources needs to be set up so that the same level of confidentiality, seclusion, and dependability is accomplished as if they were a part of a per-user dedicated infrastructure.

Additionally, in order to prevent interference with operational processes and the transmitted traffic of other services, the network that provides connectivity to computing resources needs to be protected from outside applications (e.g., by not having IP address dependencies from the rest of the other services in the network).

A logically segregated network for each tenant can be provided using a variety of techniques and technologies. Setting up an overlay network to independently transfer information between data centers or from an access network to a data center is the basic idea. Basically, the overlay network might be based at layer 3 or layer 2. The IP/MPLS (or perhaps optical) wide area network-accessible DCs' border nodes will often be connected by this overlay network.

Multiple virtual networks can then coexist over the same physical infrastructure thanks to a virtualized networking environment. Virtual Local Area Networks (VLANs) and Virtual Private Networks were the first to allow the idea of several coexisting networks (VPNs).

A VLAN is a collection of logically related hosts that are placed under a single broadcast domain. With clearly defined use cases, VLANs have developed into a widely adopted standard. The concept is that, in a more oversimplified situation, the traffic from the various tenants' VMs within the DCs is split using VLANs, which then connect to a VPN. Unfortunately, the dynamicity, flexibility, and scalability requirements, required for both virtual network configuration and proper operation, cannot be met by VLAN-based solutions. Additionally, several researchers and business vendors are working to expand and modify current network paradigms to meet the new needs brought on by virtualized use cases. Separate clients can use the same addressing scheme in different VPNs, but the IP addressing within the VPN must be unique (i.e., not duplicated). On the other hand, layer-2 VPNs (L2VPN) can be configured, functioning in base of the MAC addresses instead of the IP addresses, when simple layer-2 connectivity is needed in point-to-point or point-to-multipoint.

VPNs cannot support a network virtualization environment where dynamism, flexibility, and scalability are essential qualities because they are too inflexible. One notable example is managed network VPN (for instance,

BGP/MPLS), which is a widely used network service for businesses. The typical dynamics of cloud services are not compatible with this type of service because it was designed to operate in a relatively stable network environment, which is the case with the majority of enterprise networks in use today. Essential cloud characteristics like elasticity and self-provisioning cannot be handled by the conventional VPN approach, so those characteristics must also be extended to network resources. Quite often, expanding or reducing cloud resource capacity, or provisioning new cloud resources, requires a corresponding reconfiguration of network resources, e.g., bandwidth assigned between two data centers, whether they are in the same geographical place or not, or between the data center and the end user. In order to cope with the cloud, future network services will certainly require on-demand and self-provisioning properties.

Today the network can provide static connectivity to cloud resources, to what we call conventional networking. The next evolutionary step is to make the network elastic and adaptable according to the cloud dynamics.

Lately, it has become clear that the overlay based approach is the correct answer for achieving independency from the physical networking infrastructure. An overlay network can be created on top of an existing network, by generating logical communication links between hosts within the service domain. Overlay networks enable the design of modular networking protocols and services in which logical functions are separated from the underlying physical infrastructure.

A number of providers have worked hard to develop effective overlay systems based on various tunneling protocols. VXLAN, NVGRE, and STT are recent methods that build on overlays to achieve scalability benefits in multitenant virtual networks. Reference describes various DC connectivity methods used nowadays. VXLAN stands out among them as the most popular technology. VXLAN's fundamental idea is the encapsulation of a genuine Ethernet frame over a UDP packet exchanged between two appropriate Network Virtualization Edges (NVEs). The virtual switches to which VMs are connected for internal communication in the DC serve as NVE. The node at the DC's border will act as an NVE when communication between DCs is necessary and stitch that traffic to the inter-DC overlay network. The VXLAN Network Identifier (VNI), a 24-bit field in the VXLAN header, enables per-tenant network distinction. On top of a layer-3 overlay transport, it is therefore possible to build a virtualized end-to-end layer-2 network using VXLAN.

Overlay networks' independence from the underlying infrastructure and from one another is the main improvement they offer. This division makes it possible for independent address spaces, guarantees isolation, and enables the administration of various virtual networks by various administrators.

3.1.1.2. Network Programmability

3.1.1.2.1. Software Defined Networking (SDN)

Software Defined Networking (SDN), a model for network control that separates the control and forwarding logic and moves the traffic handling decisions from the network elements themselves to centralized software controllers, is currently the most well-liked paradigm for vendor-neutral network programmability. Conceptually, in SDN networks, the control logic is implemented on top of a so-called SDN controller, while forwarding (physical) devices have little intelligence. The controller, a logically centralized entity, is in charge of a number of responsibilities, including the creation of forwarding logic specific to a given application situation as well as the extraction and upkeep of a comprehensive picture of the network architecture and state.

SDN usually uses the Openflow protocol, which was developed at Stanford University and is now maintained by the Open Networking Foundation, for communication between controllers and network components. Openflow is now the most used SDN driving standard. The Controller can order specific rules to SDN-capable switches using OpenFlow. These rules specify how flows that fit certain criteria should be handled, including whether they should be forwarded, rerouted, changed, dropped, or QoS-shaped.

Since it can offer centralized per-flow control throughout the network and orchestrate virtualization processes, SDN opens up new views in network administration and is regarded as a significant enabler for cloud networking.

Although the OpenFlow protocol is rather low-level on its own, a number of Controller Application Programming Interfaces (APIs) have been made available to help with high-level networking application programming. The OpenFlow protocol is abstracted by these controllers to a programming language used to

write network applications. In this situation, it is simple to create management apps for cloud networking by making use of a standard set of architectural patterns, methods for querying data flows from one or more network devices, and supporting framework features.

The first widely used OpenFlow controller was the NOX controller. Initially created by Nicira and made available as open-source software. NOX soon established itself as the de facto reference design for OpenFlow controllers due to its early availability and simplicity. As a result, it has been actively used in research and feasibility studies, and it has been used to test new OpenFlow capabilities and creative controller concepts. The C programming language is used to implement NOX programs, which are referred to as modules. Because NOX is event-based, each module essentially consists of a group of callback routines that are called when particular OpenFlow protocol messages arrive. Python is supported by a NOX offshoot named POX for use in programming modules. While NOX/POX is extremely versatile it is not primarily aimed for production use, as it is not optimised for performance and stability and lacks resilience features.

Beacon, Maestro, and FloodLight are three additional controller frameworks that are designed for deployment in real-world settings. They are all Java-based and all of them use controllers. The open source foundation for Big Switch's for-profit OpenFlow controller is FloodLight.

In addition to the frameworks listed above, there are SDN management platforms that offer more services overall, making them integrated stand-alone solutions for the management of SDN infrastructures. The majority of them also make use of SDN's multi-tenancy support and network virtualization capabilities to provide these services, which are frequently referred to as "Network-as-a-Service." The following sections provide an overview of these SDN management solutions.

3.1.1.2.2. 5G Network Exposure Function (NEF)

The 3GPP 5G specification introduces a core network model that looks very different from the traditional architecture. With the aim to support fragmentation within the network and promote more dynamic 5G services, it defines an "open" core, in which all core network functions have been virtualized. This approach allows for the elimination of resource inefficiency and performance degradation associated with virtual machines and hypervisors, thereby improving the network in terms of flexibility, speed, and automation. Key enabler for this openness is the realization network programmability through standard APIs, so that higher-level service orchestrators can handle configurations for a variety of services and slices. This endeavor shapes a new and dynamic ecosystem in mobile networks from both the technology and marketing perspectives. External third parties with permission, such as industries, platform developers, and designers, may use those standard APIs for building network-aware (5G-enabled) applications, which establish a bi-directional communication with the 5GC, retrieving network statistics, but also triggering specific policies and commands to the network.

The above-mentioned exposure capability is materialized through the Service Based Architecture (SBA), adopted by the 5GC network. Indeed, the 5GC control plane NFs communicate through API-calls that define the related Service Based Interfaces (SBIs). In this context, the Network Repository Function (NRF) allows other NFs to register their services, which may subsequently be discovered by other NFs. This allows for a versatile implementation, in which each NF allows other approved NFs to access resources.

In addition, the Network Exposure Function (NEF), provides a set of northbound APIs for exposing network data and receiving management commands. More precisely, NEF provides adaptors for connecting the southbound interfaces with the SBA to an exposure layer with northbound interfaces offered to third-party developers. The overall approach is illustrated in Figure 6. In this way, NEF facilitates the secure disclosure of network resources to 3rd parties, such as network slicing, edge computing, and machine learning utilizing the 5G system, fully compliant with the innovative paradigms that underpin a wide range of services.

The functionality provided by NRF and NEF to 3rd parties, enables programmability and adaptability of the 5G connectivity services, and creates a new ecosystem where 3rd parties' developments bridge 5G exposed capabilities and service requirements/potentials from the vertical industries.

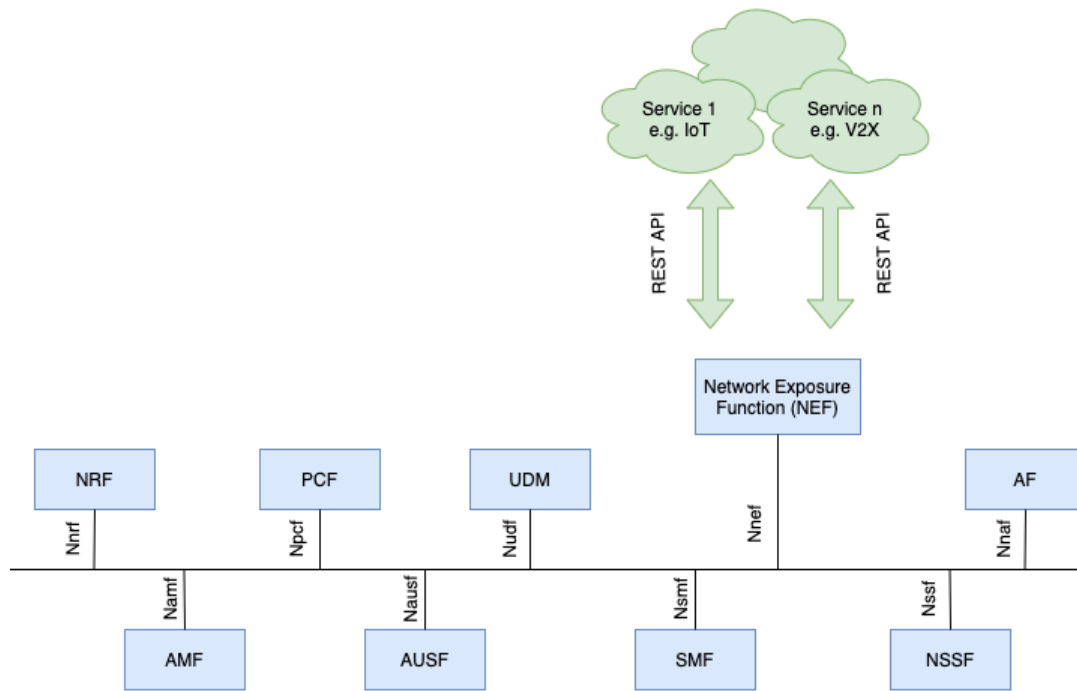


Figure 6. RESTful APIs for the Service Based Interfaces and Northbound communication

In this framework, 3GPP introduced the concept of Vertical Application Enablers (VAEs) in Rel. 16, enabling the efficient use and deployment of vertical apps over 3GPP systems. The specifications and the architecture are based on the notion of the VAE layer that interfaces with one or more Vertical apps. VAEs communicate via network-based interfaces that are well-defined and version-controlled. The focus of VAEs is to provide key capabilities, such as message distribution, service continuity, application resource management, dynamic group management and vertical app server APIs over the 5G system capabilities.

3GPP has already established the foundations to provide 5GC Network capabilities to vertical industries. The key concepts that have emerged are the Common API Framework (CAPIF) and the Service Enabler Architecture Layer (SEAL) together with NEF.

3.1.1.3. Cloud Networking

3.1.1.3.1. Network Functions Virtualization (NFV)

There is currently a significant interest in moving the network operations infrastructure of telecom operators to the Cloud as well, following the success of the cloud computing/storage paradigm where computation and data are moved from end-user devices to dedicated servers. By separating the hardware and software of these network pieces and replacing the former with commercial off-the-shelf (COTS) devices, this program aims to lower the CapEx and OpEx of such infrastructure. A virtualized network function (VNF) can be implemented across hypervisors in this new design, which are completely transparent to the actual hardware below. Because the hypervisor offers a common interface to access virtual compute, storage, and network aspects, a VNF can be executed over any hardware platform compatible with the hypervisor in this fashion.

Since a few years ago, both network administrators and equipment producers have been developing technologies related to network virtualization, with a focus on the concepts of running network operations on general-purpose servers and cloud infrastructures, which are largely motivated by the impact of cloud technologies on all areas of IT. The industry advancement in this area was facilitated by encouraging results on the performance of these solutions for actual network workloads at the start of this decade.

Industry Specification Groups (ISGs) are used by the European Technical Standards Institute (ETSI) to offer a fast route for the development of industry fora on certain themes. The Network Functions Virtualization (NFV) ISG is a representative example of one of these organizations. It aims to address the issues that network operators face and that are brought about by the ever-increasing number of network functions that are implemented in specialized appliances, including the need to find room and power for them, the requirement for specialized

device handlers, the short life-cycle, etc. By utilizing industry-standard IT virtualization technologies, NFV seeks to address this issue by consolidating as many network operations onto equipment commonly found in modern datacenters. NFV is complementary to Software Defined Networking (SDN): while network functions can be virtualized without the need of an underlying SDN infrastructure, both are mutually beneficial.

NFV is a technique (or a group of technologies) designed to create network infrastructure services using current cloud infrastructures in the same way that IT services are created. In order to accomplish the real network functions, participating software components are anticipated to access a common virtualization interface through a homogenous supporting infrastructure that provides compute, storage, and networking mechanisms. The dual function of network facilities must be noted. To facilitate the interconnection of the components (hardware and software) needed by the software modules implementing the second, upper layer of network operations running on the infrastructure, there is a layer of homogenous, virtualized network methods.

The architectural framework, presented in, provides the blueprint for vendors to implement NFV compatible products and is made of a series of building blocks vendors can choose from. The NFV Architecture depicted in Figure 7 is comprised of four main functional elements [SNIM-1]:

- **The Virtual network function (VNF)** layer virtualizes a certain NF, that operates independently of others. A particular VNF can run on one or more VMs and it can be divided into several sub-functions called VNF Components (VNFCs). VNFCs monitoring is performed using Elemental Management Systems (EMSs). Automation of the operational processes is feasible and results in improvement of the efficiency and reduction of the OPEX costs.
- **The NFV infrastructure (NFVI)** is comprised of all the hardware and software required to deploy, operate, and monitor the VNFs. Particularly, NFVI includes a virtualization layer necessary for abstracting the hardware resources (processing, storage, and network connectivity) to ensure independence of the VNF software from the physical resources. The virtualization layer is usually composed of virtual server (e.g. Xen [SNIM-2], Linux-KVM [SNIM-3], Dell-VMware [SNIM-4], etc.) and network (e.g., VXLANs [SNIM-5], NVGRE [SNIM-6], OpenFlow, etc.) hypervisors. The NFVI Point of Presence (NFVI-PoP) defines a location for network function deployments as one or many VNFs.
- **NFV management and orchestration (MANO)** is comprised of three components:
 - *The virtualized infrastructure manager (VIM)*, which has the responsibility to manage and control VNFs interaction with physical resources under its supervision (e.g. resource (de)allocation, inventory),
 - *The VNF Manager (VNFM)*, with the responsibility to manage VNF life-cycle (e.g. link initialization, suspension, and termination),
 - *The NFV Orchestrator (NFVO)*, with the responsibility to realize network services on NFVI and to additionally monitor operations of the NFVI collecting information regarding operations and performance management.
- **Operations support systems and business support systems (OSS/BSS)** element comprises the legacy management systems and assists MANO in the execution of network policies. The two systems (OSS and BSS) can be operated together by telecommunications service providers or operators, either automatically or manually to support a range of telecommunication services.

The overall framework produces a practical and operational “virtual” network. At the bottom lies the **Virtualized Infrastructure Manager (VIM)** which provides the proper functionalities to control and manage the underlying infrastructure components, including storage, computational and network resources. Ultimately, VIM interconnects Virtualized Network Functions (VNFs) with the physical resources, acting like a hypervisor on virtualization framework. VIM is connected with the **VNF manager (VNFM)** through Vi-Vnfm interface. VNFM controls and manages the lifecycle (e.g. instantiation, update, termination) of VNF instances. ETSI specification assumes that each VNF instance has an associated VNFM; however, a VNFM may correspond to a single or multiple instances. **NFV orchestrator (NFVO)** has two main responsibilities, the orchestration of Network functions Virtualization infrastructure (NFVI) resources and the lifecycle management of Network Services (NSs). Network services are compositions of individual interconnected VNFs. Generally, NFVO

brings on new network services and VNFs and provides a global resources management. In addition, it is responsible for validation and authorization of NFVI resource requests.

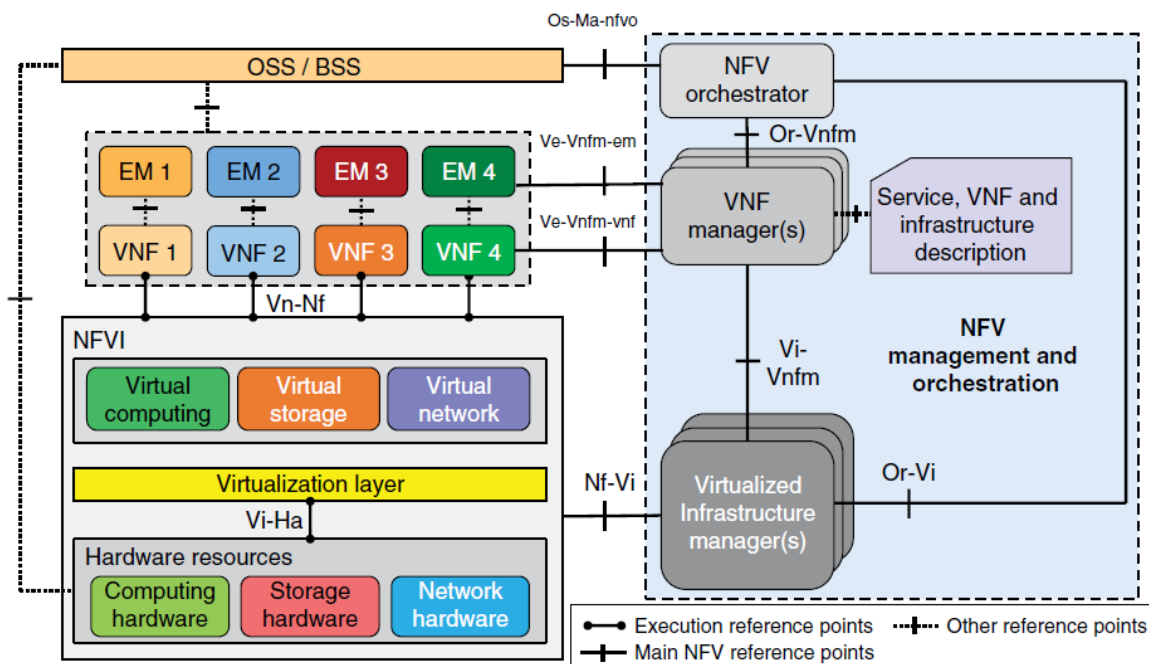


Figure 7. ETSI NFV architecture

NFVI is a key component whereas the totality of hardware/software is accumulated and on which virtual networks are built. In other words, it is the proper environment in which VNFs are deployed. NFVI is categorized into three components, the hardware resources, virtualization layer and the virtual resources. The hardware resources of NFVI are widely available, low-cost and primarily standardized (i.e., commodity hardware). Virtualization layer is mainly a hypervisor that ensures VNFs are decoupled from the underlying hardware. Virtual resources are practically VMs in which VNFs are deployed.

The NFVI may include partially virtualized network functions [SNIM-7] such as hardware load balancers, DSL Access Multiplexers, Wi-Fi access points etc. The other part of the functionality may depend on vendor design choices. At this point we should mention that SDN and NFV specifications aim to provide interoperability among vendors; therefore, some vendors may deviate from the specs to differentiate their products. OSS and BSS are expected to have information exchanges with the NFV MANO architectural framework (through OS-Ma-nfvo interface) in order to provide management and orchestration of legacy system.

The main benefits of deploying network services as virtual functions are: (1) flexibility in the allocation of network functions in general-purpose hardware; (2) rapid implementation and deployment of new network services; (3) support of multiple versions of service and multi-tenancy scenarios; (4) reduction in capital expenditure (CAPEX) by managing energy usage efficiently; (5) automation of the operational processes, thus improving efficiency and reducing operational expenditure (OPEX) costs.

While this was probably an initial guess in many cases, there are three essential aspects that distinguish NFV from the direct application of cloud technologies to provide network infrastructure services, and therefore require going beyond carrier clouds to implement NFV.

First of all, the workloads that NFV implies are entirely different from the workloads that the existing cloud practice takes into account. Direct I/O and memory operations, as opposed to direct processing or storage access, are heavily relied upon by VNFs. Additionally, this affects the portability of VNF instances throughout the cloud architecture, which is much more relevant, in addition to the performance of VNF when deployed directly using "traditional cloud" mechanisms. Improved cloud orchestrators, hypervisors, kernels, and even hardware drivers are required to support more precise placement policies, give better control over direct memory communication between software instances, and bypass the virtualization layer for direct I/O to network interfaces in order to properly achieve performance and portability goals.

Network services also need to adjust to the shape of the network. While most network infrastructure services are middle-points (such as a router or firewall) and many of them are subject to strict latency requirements and/or similar constraints, traditional cloud apps are endpoints in a connection (the prototypical web server in many cases). This indicates that infrastructures and VM placement techniques must adjust to the network's architecture and support both highly centralized and consolidated datacenters when they are applicable and their economies of scale can be utilized, as well as far more decentralized schemas. Supporting both types of deployments while also being able to smoothly integrate them is crucial in this situation.

The supporting infrastructure that is already present in the current clouds and the upper network service layer offered by VNFs and their composition into services are the two networks that we are dealing with when it comes to the orchestration and management of the resources. Upper network services may need to actively modify the underlying network infrastructure in order to ensure performance, going much beyond the typical northbound interfaces exposed by the SDN controllers now being used in cloud datacenters.

Undoubtedly, cloud computing is a fundamental NFV enabler and the idea of NFV itself. NFV must make use of the technologies now used in cloud computing. These solutions rely on hypervisor-based hardware virtualization processes and virtual Ethernet switches to move traffic between virtual computers and physical interfaces (though other possible virtualization mechanisms could be applicable, the current focus of the NFV community is on these techniques). Additionally, current cloud approaches offer ways to improve resource availability and usage through orchestration and management mechanisms. These mechanisms are applicable to the automatic instantiation of VNFs, resource management, re-initialization of failed VMs, creation of VM state snapshots, migration of VMs, etc.

3.1.1.3.2. Cloud-Native Network Function

Virtual network functions (VNFs) started as the virtualization of network hardware. VNFs had a one-to-one correspondence of hardware to virtualized hardware. Still the porting of software from propriety systems to virtualized machines as monolithic network functions, establishes some difficulties regarding their agility in terms of scalability, resilience and quick application evolution among others. Design patterns, of what came to be known as **cloud native software** (CNF), have emerged over the last few years providing complex services decomposition into an architecture of loosely coupled, stateless components which should be able to communicate over language-agnostic APIs. This pattern provides a vision of the cloud as an entirely new kind of distributed computing environment, that opened up exciting possibilities for new application architectures, instead of writing software to run on dedicated servers and then deploying it on virtual machines in the cloud. This new approach, if appropriately implemented, may offer increased flexibility, scalability, reliability, and portability.

From an implementation point of view the distinguishing feature of the cloud-native approach is that it uses containers rather than VMs. This allows network functions to be provided as a software package, with all dependencies necessary to run it included, while sharing access to the operating system and other server resources. This enables an easy transfer and placement of the contained components among completely different environments (e.g. production, development) and even among clouds while at the same time they retain their full functionality without the need of any adaptations or modifications.

By design cloud-native network functions (CNFs) are implemented to run inside containers. This containerization of network architecture components makes it possible to run a variety of services on the same cluster and more easily on-board already decomposed applications, while dynamically directing network traffic to the correct pods.

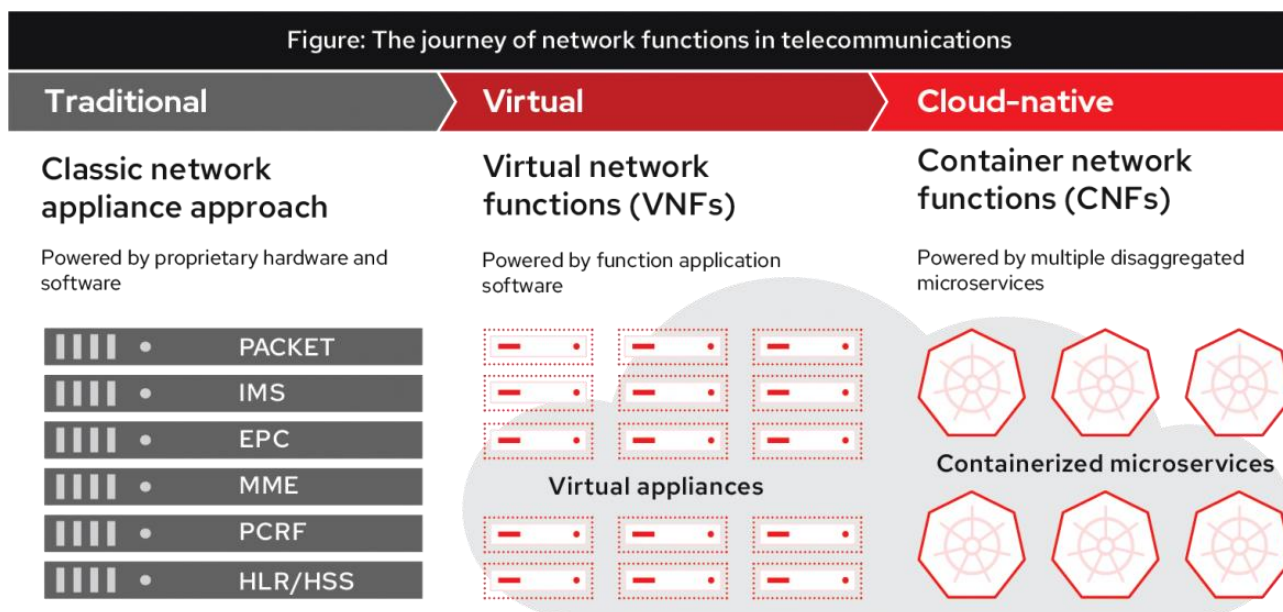


Figure 8. Network functions evolution [SNIM-8]

Just the process of containerizing network functions does not guarantee a cloud native approach. In fact CNCF has published a “trail map” [SNIM-9] that provides guidance on best practices for cloud native application development.

When applying this to cloud native network functions, we end up having to implement the network function just like any other cloud native application. A summary of this “trail map” is as follows:

1. Containerization of applications and dependencies with an emphasis on splitting it and deploying a set of a coarse-grained set of microservices.
2. Set up of a CI/CD pipeline with stateless and declarative configuration for the service or application, so that changes may lead to an automated new service build and deployment.
3. Orchestrator support and deployment for the services lifecycle management.
4. Embody network function monitoring, tracking and logging, by deploying telemetry facilities for metrics and tracing.
5. Ensure, service discovery, which will allow network service to be discovered by other consumers inside or even outside of the cluster.
6. Facilitate declarative configuration, by outlining the importance of policies, especially network and security policies, as being applicable and supported through the service.
7. Distributed storage, whenever stateful workloads are used, ensuring thus compatibility with cloud native environments.
8. Cloud native messaging like pub/sub, request/reply systems.
9. Efficient runtime delivery and software distribution (e.g. accessible registries)

The first three steps of the above described process are imperative in order to declare the architecture as cloud native. The enforcement of these techniques enable loosely coupled systems that are resilient, manageable, and observable. Combined with robust automation, they allow engineers to make high-impact changes frequently and predictably with minimal toil [SNIM-10].

Key features of cloud native application architecture include patterns like these referred below.

- **Stateless processing**, which dictates the deployment of a highly scalable and very fault-tolerant storage system in which to keep all of our application state.
- **Microservices** pattern, which provides composability and reusability by developing modular components each of which performs a very specific task and exposes a well documented API, providing thus technology heterogeneity, efficient scaling and ease of development and deployment.

- **Containers** which approach virtualization from an aspect that happens to be particularly well-suited to cloud native applications with the use of Linux Containers. Containers provide a number of advantages which include low overhead, minimized startup latency, reduced maintenance, ease of deployment and high portability.
- **Design for automation**, as cloud native applications comprise a large amount of different software components. Proper orchestration is required for the deployment and management of the various components. For this reason, orchestrators have emerged to help manage microservices. Orchestrators are in charge of the scheduling, starting, stopping, and monitoring the lifecycle of the microservices.
- **Declarative configuration** allows for the whole system to be self-healing because it makes it easier to read and respond to what the system should look like. The system can then be made to continuously correct itself.

CNF benefits to be considered are the provision of better resource efficiency by running more services on the same server (using the native structure of microservices and concept of containerization), resiliency and higher availability, as microservices are spread over multiple servers and machines and the processing load is shared, higher development velocity for scaling the network using Kubernetes orchestrator, less downtime in the network using rolling upgrades of microservices.

3.1.1.3.1. 5G Network App (5G-NetApp)

Considering the 5G openness capabilities, materialized through APIs, as described above, in this section the concept of the Network Application (NetApp) is defined. More precisely, in the context of EVOLVED-5G as NetApp is defined a software piece that interacts with the control plane of a mobile network by consuming exposed APIs (e.g., Northbound APIs of 5G core and/or MEC APIs) in a standardized and trusted way (i.e., for a 5G network a NetApp should be CAPIF compliant;**Error! No se encuentra el origen de la referencia.**) to compose services for the vertical industries.

A NetApp shall provide services to vertical applications either as an integral part of the vertical application or by exposing APIs, which are referred to as business APIs. In this context, vertical industries will be able to develop NetApps that compose new services by consuming 3GPP APIs as well as other telco assets (referring to business support system – BSS APIs, e.g., service orchestration APIs).

For example, authors in [SNIM-11]**Error! No se encuentra el origen de la referencia.**, proposed a framework that leverages NEF APIs (i.e., TrafficInfluence API to influence data-path configurations and MonitoringEvent API to retrieve location information) to plan where to place Video on Demand (VoD) content. The framework distributes segments of the full video to MEC caches, but only a portion will effectively be consumed while the user traverses MEC's coverage area, minimizing access time (i.e., low latency) and optimizing traffic load on the core network. The components that carry out this activity in the proposed framework, could be considered as NetApps. Note that, this potential NetApp not only receives information from the 5G Core but utilizes these data to perform a more intelligent task. Machine learning algorithms can be applied on the framework to predict where to place the segments. Therefore, considering the way that the services are provided to verticals, the NetApps can be classified to:

- **Standalone NetApp.** A standalone NetApp provides complete services to one or more vertical industries, either directly or through its integration to a vertical application. A NetApp that is integrated into a vertical application, enhances the functionality of the application by adding network management and monitoring capabilities exposed by the 5G network.
- **Non-Standalone NetApp.** NetApp that operates as a wrapper of Northbound APIs to expose services through Business APIs. It is an auxiliary non-standalone software piece (in the sense that it becomes functional when its business APIs are consumed by an app). A Non-Standalone NetApp allows vertical applications to be developed/upgraded (and take advantage of the 5G exposure capabilities) without changing integral parts of their software, i.e., only by consuming the business APIs.

The two types of NetApps are presented in Figure 9:

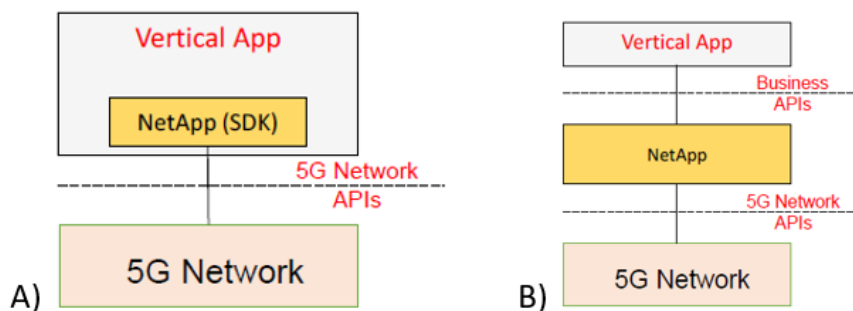


Figure 9. Third-party Standalone (A) and Non-standalone (B) NetApp representation

The NetApp ecosystem is something more than the introduction of new vertical applications that have 5G-interaction capabilities; it responds to the request for a separated middleware layer that will simplify the implementation and deployment of vertical systems at large scale (considering also the adaptation needed for Non-Public 5G Network – 5G NPN deployments). This is the same request that triggered the development of Vertical Application Enablers (VAE) by 3GPP SA6. NetApps can also be categorized by the level of interaction and trust with the Mobile Network Operator (MNO):

- **Third-party NetApp.** NetApp that resides at a trusted third-party domain. A third-party NetApp consumes Northbound APIs and, also, supports trust mechanisms and security policies defined by the network for the verticals.
- **Operator NetApp.** NetApps that reside at the operator domain, considering mainly Non-Public Network (NPN) deployments, and, potentially, can have further access to 5G network capabilities, beyond those provided through the Northbound APIs (e.g., vertical specific functionality at the OSS for slice management) and those available in a third-party NetApp. In that case, the NetApp may interact directly with the 5GC NFs.

Considering the 5G SBA, a NetApp can be an Application Function (AF) that assists the vertical server client to communicate with the 5GC network (i.e., control plane) and utilize its capabilities to enable network-aware applications. Note that, a NetApp is part of the VAS as defined by 3GPP SA6, thus a NetApp is instantiated during the development time of a VAS.

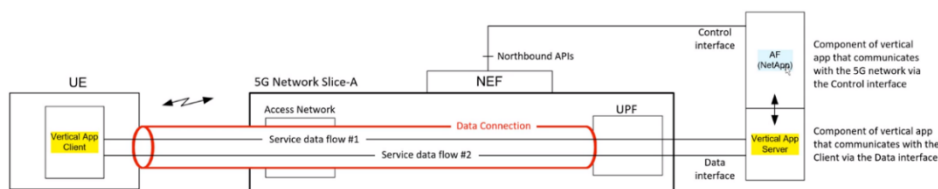


Figure 10. NetApp's interaction with the data and control plane when a Vertical application is provided

3.1.2. Resource orchestration approaches

Resource orchestration, from the perspective of aerOS, can be conceived as the distribution of computing power based on the deployment of services. In this sense, it has been powered by the recent advances in network function virtualization (NFV) technologies and software defined networking (SDN) technologies. Although the smart networking has been thoroughly reviewed in the previous section, it is key to reinforce some ideas related to SDN and NFV in order to understand how resources orchestration can behave in the continuum. In this section, these advances are analysed from the point of view of resources distribution across the computing elements.

It supports seamless and elastic service deployment for verticals while efficiently reusing the available resources, and thus, reducing incurred costs and consumed energy. It is becoming more and more as an inevitable solution that enables fast service delivery, reduces human intervention and ensures a good and consistent Quality of Service for the deployed services.

Effectively, with NFV, Virtual Network Functions (VNFs) deployable inside telcos datacenters decrease the operating cost and improve performance. VNFs are easy to deploy, upgrade, and scale up or down. They are also more fault-tolerant than their counterpart functions built on dedicated hardware. Moreover, SDN technologies can be integrated with NFV to provide easy and remote configuration of network equipment. This allows better management of the infrastructure with the concept of network slicing [ROA-1]. Currently, ETSI and IETF are the maintainers of both NFV and SDN specifications, respectively. ETSI has defined a NFV reference architectural framework illustrated in Figure 11. SDN resources, controllers and applications can be placed at different locations in this architecture as reported in [ROA-2]. ETSI-NFV defines a network service architecturally as a forwarding graph of Network Functions (NFs) interconnected by supported network infrastructure as illustrated in Figure 12.

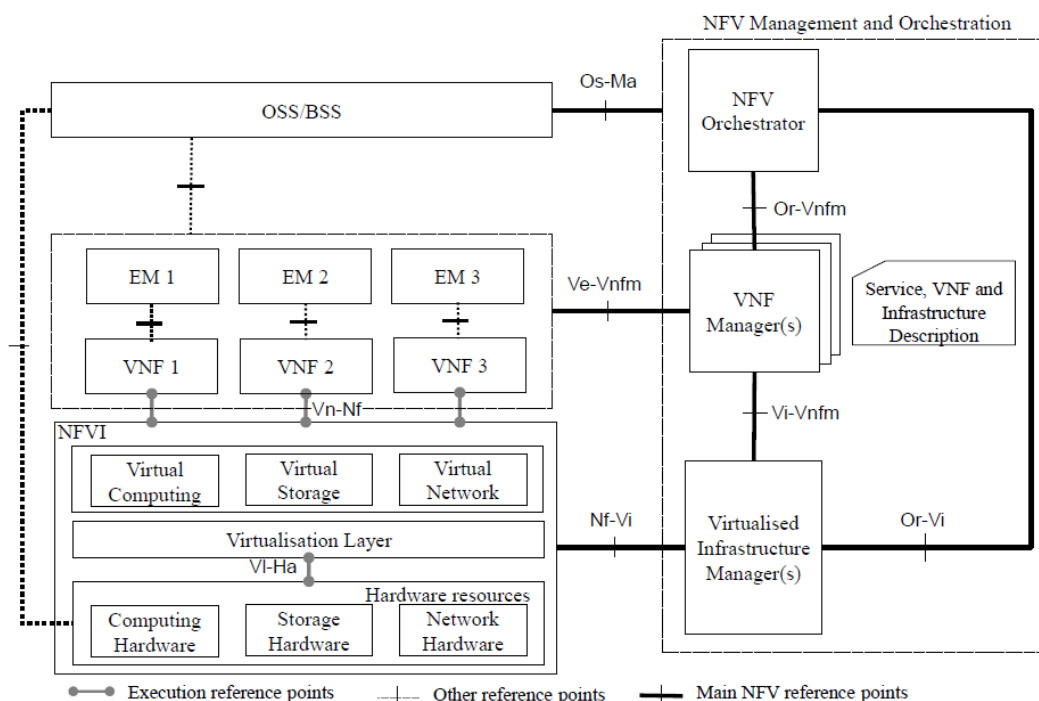


Figure 11. NFV Reference Architectural Framework [ROA-3]

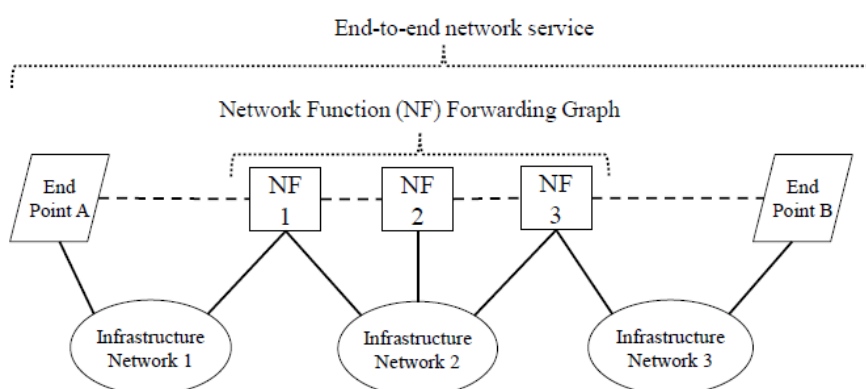


Figure 12. Graph representation of an end-to-end network service [ROA-3]

In a single administrative domain, the NFV Management and Orchestration (MANO) orchestrates the different network services and manages the VNFs and the underlying virtualized infrastructure. NFV descriptors based on TOSCA specification are used for the VNF deployments [ROA-4].

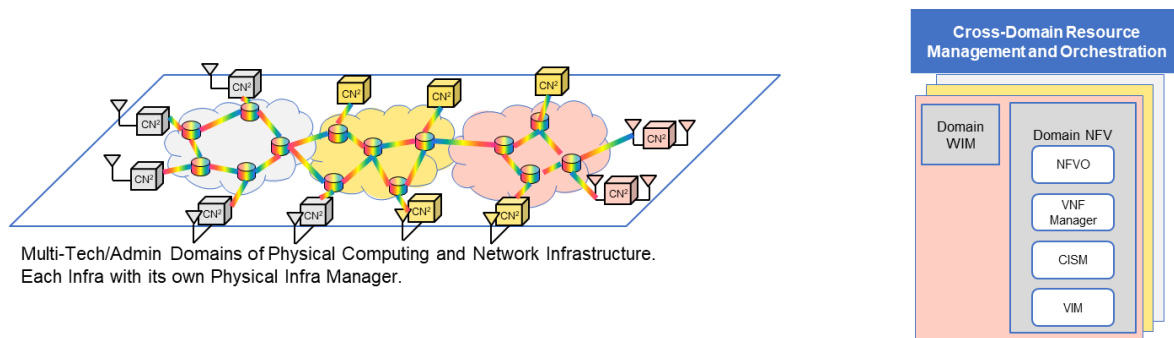


Figure 13. Resource management and orchestration across multi-technological and administrative domains

Support for multiple administrative domains is also important (Figure 13). ETSI has defined the functional requirements, interfaces and operations to support the provision of network services across multiple administrative domains. They are based on the interactions between NFV Orchestrators in different administrative domains (supported over the Or-Or reference point) [ROA-5]. ETSI has also reported the different architectural options to support multiple administrative domains for NFV MANO [ROA-6]. Effectively, with the advent of 5G and its different traffic types and supported service, orchestration encompasses now the Radio Access Network, edges and clouds. Thus, one of the major challenges for Service Orchestration is to efficiently orchestrate services in this heterogeneous continuum of resources federation (Authority, Technology, Location).

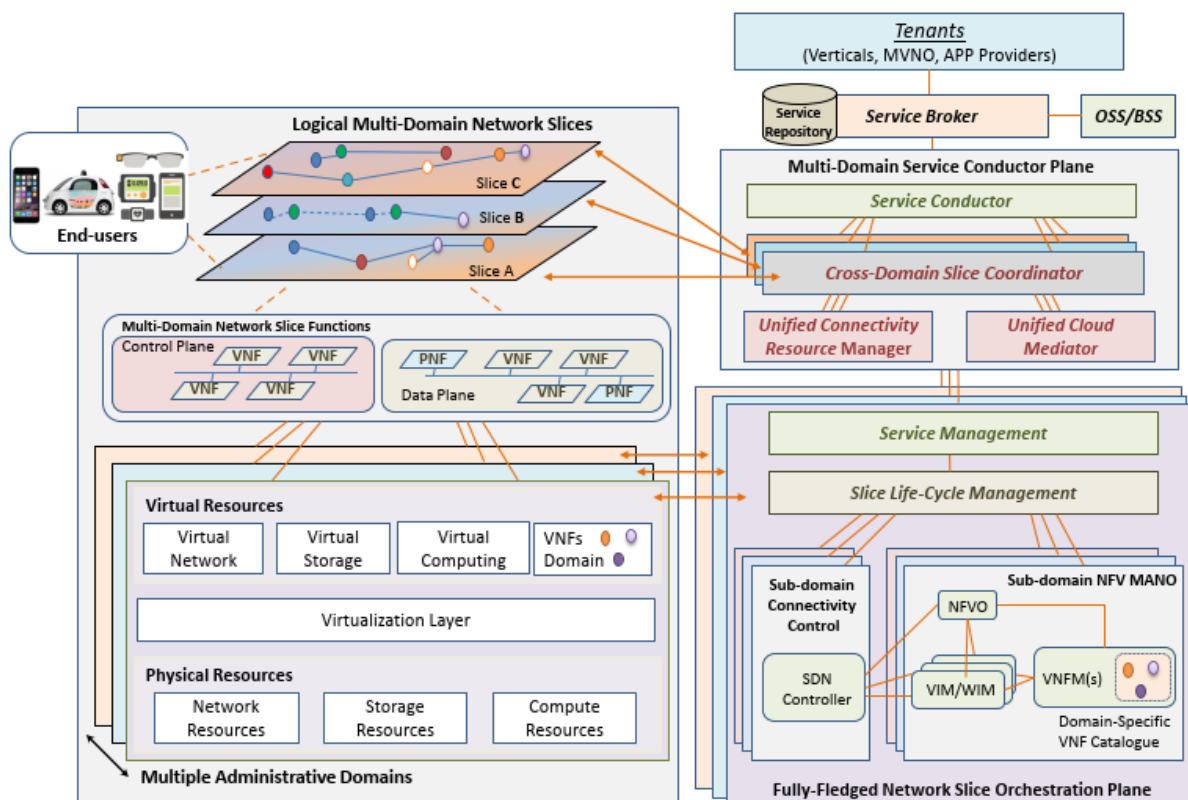


Figure 14. An architecture for multi-domain resource orchestration for network slicing [ROA-8]

As opposed to the single domain orchestration where the orchestrator has full control over the resources, the multi-domain orchestration requires some sort of coordination between the different domains. Where in the former case, the resource orchestrators can be used to optimize against one or a set of objectives (e.g., QoS, Cost, and Energy) [ROA-7], in the latter case, the coordination between the domains should find some sort of equilibrium of the objectives inside each domain and across the domains while ensuring the QoS of the E2E service. In this vein, different solutions have been proposed. For example, the work in [ROA-8] proposes an entire framework that incorporates SDN and NFV components to the basic 3GPP network slice management.

The framework consists of four major strata: multi-domain service conductor stratum, domain-specific fully-fledged orchestration stratum, sub-domain management and orchestration (MANO) and connectivity stratum, and logical multidomain slice instance stratum (Figure 14). Each of these strata has its own fundamental operational specifics for instantiating and managing the resulting federated network slices.

There are many concepts and solutions that permit the management of end to end (E2E) services, above all in a zero-touch manner [ROA-9][ROA-10]. However, their orchestration approaches may differ according to at which extent their decision engines are centralized or distributed. Indeed, centralized orchestrators have a complete knowledge on the underlying infrastructure. Such orchestrators offer relatively short convergence times, but they have limited scalability and they represent a single point of failure. While for the distributed orchestrators, they are robust and have good scalability, but the convergence times are longer due to peer-to-peer negotiations that require complex interaction [ROA-11]. Due to the ever-growing virtualization, lately, more research is targeted towards hierarchical orchestration. In such orchestration, each domain orchestrator manages its own resources and services and they are coordinated by higher-level orchestrators that can be also coordinated by even higher-level orchestrators [ROA-12].

The growing complexity in orchestrating services calls towards an automated orchestration and management of the service. As stated earlier, NFV and SDN have played a crucial role in the softwarisation of the network and its management. However, the latter still remains traditional, in a sense that it is based on pre-defined policies that are static over relatively long periods of time. There is still a long way to go till a 100% closed-loop autonomous management of networks in a zero-touch fashion [ROA-9][ROA-13]. In this vein, different initiatives have been kicked off, with ETSI ZSM ISG (Zero Touch network and Service Management Industry Specification Group) [ROA-14], ETSI ENI ISG (Experiential Network Intelligence), and TMF's Zero-touch Orchestration, Operations and Management (ZOOM) being the most noticeable ones. All of these follow the concept of closed-loop automation (CLA) in order to realize zero-touch orchestration and management. The CLA is based on the two most notable paradigms for closed control loop which are OODA (Observe, Orient, Decide, Act) and MAPE (Monitor, Analyse, Plan, Execute). Such as seen above, the CLAs of the domains can be run independently (distributed), hierarchically or they can be federated. Hierarchical CLAs form a tree of CLAs where decisions and results of a CLA are communicated to a supervisory CLA. Distributed CLAs are physically distributed and communicate and negotiate with each other. For federated CLAs, an agent is used to govern a set of CLAs that have the same goal but where each one of them uses its own data and where their decisions are aggregated and exchanged among them.

Artificial Intelligence (AI), supported by Machine Learning (ML) and Big Data analytics techniques, is envisioned as a key enabler to realize Zero-touch orchestration and management [ROA-15][ROA-16]. Such systems would be of no value without a robust and high-performance algorithmic framework that governs, in an autonomous fashion, all operational processes and tasks, starting from the planning and design of the network, towards its deployment, resource provisioning, monitoring and optimization. The success of these operations hinges largely on the choice of the AI/ML model and its interpretability, which, in turn, depends heavily on the availability of high-quality data. Furthermore, the next generation zero-touch management system should be able to make highly accurate decision making, above all in real time or near real-time. This is particularly of vital importance for next generation networks promising the further support of ultra-low latency and ultra-reliable communications. On the other hand, high accuracy of AI/ML techniques comes at the cost of high demand for computation resources. Hereby, solutions to optimize and accelerate the execution of AI/ML techniques, without any loss of accuracy and whilst keeping their complexity and computation needs within an agreeable budget, become needed. The zero-touch management of the network services, along with the relevant resource orchestration, should be carried out in an end-to-end style, considering and efficiently coordinating all possible synergies among the different segments of the different domains (i.e., radio access network, core network, transport network, edge/cloud, etc). In this context, suitable APIs among the different network segments, along with the supporting mechanism, should be designed not only to facilitate the data sharing, but also to ensure safe shared learning collaboration among the segments of the same mobile network and across multiple mobile networks administrated by different operators.

Automated orchestration systems face many challenges that need to be addressed to realise CLA in next generation networks. Translating vertical or service requirement in order to setup a slice is of utmost importance. Indeed, when slice customers have a unique set of QoS requirements, the resource orchestrator should map the high-level QoS requirements into the appropriate set of VNFs characterized by their compute, memory and

storage requirements, their locations, level of isolation from other slice, and also the links requirements between the VNFs. Currently, the resources allocated to VNFs are handcrafted by the network operators which lead to resources overprovisioning. Therefore, data-driven resource orchestration is needed in order to allocate the right amount of resources for each service. Such a mechanism should be supported by powerful predictive algorithms.

3.1.3. APIs, monitoring and communication services for the continuum

This section reports about one of the four axis of aerOS orchestration: the services themselves, understood as the software that performs an action oriented to a business, operative or functional goal. It also reports about how those services are exposed and interacted across the continuum (as far as current approaches propose).

3.1.3.1. Service Orchestration

Service orchestration is the distribution of services on the nodes of a network. Services can be applications, micro-services, or containerized environments which can be stand-alone or communicate with each other. The service orchestration consists of objectives e.g., lowest latency or lowest energy consumption and (physical) constraints e.g., resource consumption or device capabilities. In the literature, finding a feasible service orchestration is known as task scheduling, allocation or offloading problem depending on whether the focus is on time-dependent, network-specific, or user-equipment-centric orchestration. Thus, we use the word task interchangeable to service. The orchestration problem is known to be NP-hard [SO-4] and can be solved by optimal and heuristic approaches.

An orchestration is optimal if it adheres to its constraints and returns the best value for its objective. Optimal solutions can be found by using (Mixed) Integer Linear Programming [SO-4, SO-7, SO-3]. After defining the model as an ILP problem, we can use a solver like PuLP [SO-1] and IBM CPLEX [SO-2] to find the optimal orchestration [SO-3]. Cardellini et al. [SO-4] define an ILP as extendable framework for optimal data stream processing application placement. They show that this framework can be easily extended with additional QoS constraints like e.g., bandwidth consumption, inter-node traffic and elastic energy of the network. Seeger et al. [SO-7] extend this framework with the focus on reducing the overall energy consumption as objective. As identified in Seeger et al. [SO-7] and Buschmann et al. [SO-3], solving these ILPs scales poorly. For example, finding a solution takes over two weeks for over 28 tasks [SO-3]. As a result, this approach may be infeasible for dynamically changing edge networks.

Heuristics and meta-heuristics approximate the optimal solution and avoid the poor scalability of optimal approaches [SO-7, SO-3]. Seeger et al. [SO-7] propose a heuristic by using an ILP model with an approximation of the network energy consumption; thus, decreasing the complexity and scaling linearly. Other orchestration approaches use population-based meta-heuristics like the Genetic Algorithm (Skarlat et al. [SO-8]), Particle Swarm Optimization (You et al. [SO-9], Buschmann et al. [SO-3]).

Currently, the most promising algorithms in terms of time and resource consumption are based on machine learning. Gao et al. [SO-5] propose a Deep Reinforcement Learning (DRL) approach to offload workflows consisting of one or more tasks on edge servers and user equipment. They minimize the energy consumption and completion time of the workflows with a multi-agents deep deterministic policy gradient algorithm. This algorithm yields the best values and terminates as fastest in comparison to random offloading and DQN-based offloading. In the context of the IntelliIoT project [SO-6], Buschmann et al. [SO-3] propose and analyze a DRL approach for task allocation which outperforms the ILP heuristic of Seeger et al. [SO-7] and the PSO approach for problems with increasing complexity.

3.1.3.2. Application Programming Interfaces

“The dynamic generative growth of a digital ecosystem is what makes digital innovation unique” [SO-11].

APIs have a big impact on evolution on software ecosystems [SO-10]. It is essential for the growth of an ecosystem that contributors need to constantly adapt to changes of the underlying APIs. Continuous change and innovation of technology requires that API interfaces need to adjust as well. Fast-evolving APIs are used more by clients than slow evolving APIs, but the average time taken to adopt new versions is longer for fast evolving APIs [1]. This results in a dilemma for API design philosophy. For clients consuming a given API, updating to

a new API version can be costly and off-putting. On the other hand, the API developer needs to be certain that the evolution of the API is consistent, so that it covers existing or future use cases while causing minimal incompatibility with older versions for the migration effort on the client side [SO-11].

Defining the boundaries of an ecosystem is done through the API design that allows third-party developers to create add-on products for a given ecosystem [SO-12]. Based on data collected on Wordpress.org [SO-13], the co-evolution of APIs between creators and contributors for a digital ecosystem is crucial. Existing pure digital ecosystems like Wordpress give insights into how an ecosystem can successfully grow. In 2012, Wordpress featured 443 unique APIs from which only 103 originated solely from the initial platform founder. This indicates a thriving ecosystem that has grown outside of its initial design boundaries. From this, the following hypotheses emerge that APIs as the core of an ecosystem are more influential than individual components offered by a focal platform system for the growth of a digital ecosystem.

With the relevance of API stated, it is equally important to look at further influences introduced by managing and growing API developments. APIs have various impacts on the development of ecosystems from a non-technical standpoint. Based on three ecosystems of Siemens, major challenges for business, architecture, process, and organization perspectives can be identified [SO-14]. These challenges lead back to the API management for ecosystems. For the business aspect, it is important to find the optimal speed of innovation for all partners. For the architectural challenges, the management of API dependencies is crucial to the further development and maintenance of the ecosystem. Furthermore, process and organizational challenges summarize outside influences concerning APIs for an ecosystem. Concerningly, an API is traditionally observed as something sole technical from an outside standpoint. However, it must be realized that many interconnections between API development and other concerns of software development exist. By respecting these circumstances, the growth of an ecosystem can be further enabled.

With the relevance of API development stated, it is essential to investigate the further implications of API in an ecosystem related perspective. As shown in the WordPress [SO-12] ecosystem development, it is crucial to build a network of contributors outside of the core of an ecosystem. As seen in Figure 1, an ecosystem can be represented by the metropolis model. The kernel or core of the ecosystem is crucial to define how other entities can interact with the ecosystem. Therefore, the API of the kernel needs to be designed in a very flexible way to allow all kinds of prosumers and customers to interact with the ecosystem.

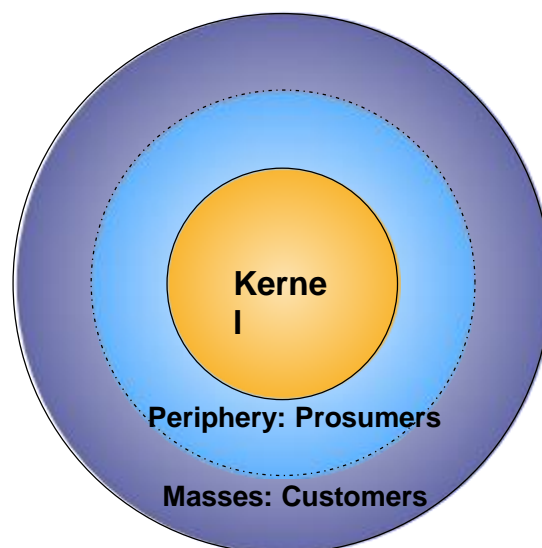


Figure 15. Metropolis model structure [SO-16]

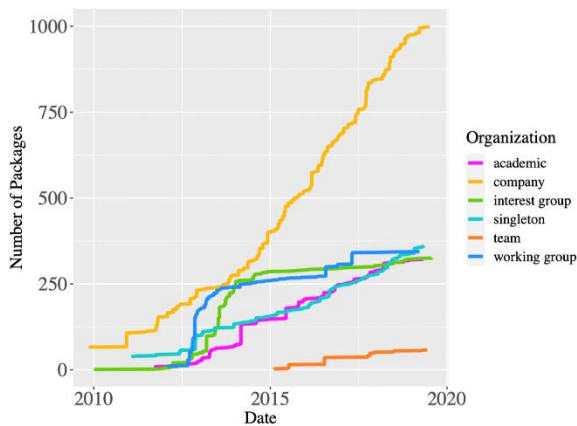


Figure 16. ROS contributors [SO-21]

With multiple domains across all kinds of fields, the existence and inclusion of mashup ecosystems with open kernels and closed kernels is very relevant. Ecosystems like ROS try to reach into multiple closed source ecosystems in order to provide a prosperous ecosystem in the domain of robotics. However, thousands of individuals would be needed to create such one-of-a-kind ecosystem as extensive as ROS from scratch [SO-17]. Especially, to reach a critical mass of prosumers like companies to support the effort of an open ecosystem, as seen in Figure 2.

By supporting the API given by the kernel of the ecosystem, prosumers are enriching the ecosystem but also personally gain attraction in form of accessibility to the rest of the ecosystem.

In the world of open-source ecosystems, ROS stands out from other communities from other domains as seen in the previous figure. Ecosystems or communities can be seen with different kind of characteristics. The Linux Kernel is for example a very protective community. This behavior has enabled the Linux community to provide a very robust outcome but change and therefore innovation is hard to achieve. Other communities like NodeJS or ReactJS see very fast grow in the web business but fail to bring consistency. Analyzing these different PR governances by [SO-18] shows what different characteristics are needed to have a thriving ecosystem that combines a multitude of APIs to a heterogenous standardization effort in the case of ROS for the domain of robotics.

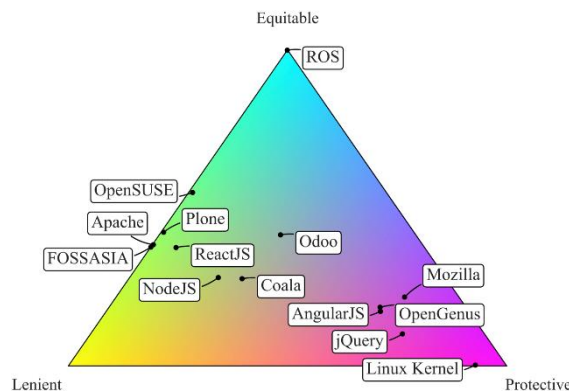


Figure 17. ROS Ecosystem PR governance [SO-18]

Still, with all aspects of the ROS ecosystem considering, the ROS ecosystem currently is tailored to its specific domain. For end users, APIs and ecosystems play a secondary role. Users are primarily facing technological problems from a use-case driven standpoint. Use-cases are however not exclusively pinned down to a single domain. Rather to the opposite, innovation driven use-cases tend to have their core differentiation in the combination of different domain solutions in order to address new markets.

Following, the state of API design proposals trying to combine multiple domains are rather tenuous and have slow momentum of the sparse community supporting such new ecosystems [SO-19]. Revisiting the standpoint of the end user, combining different APIs from multiple ecosystems in one system to support a dedicated use-case is currently not addressed during the design of conventional ecosystems.

In the era of the Web 1.0, APIs were designed to enable all kinds of users to take full advantage of all the information provided [SO-20]. With the growth of the Web 2.0 and following platforms, the overall available and defining APIs of the Web 2.0 altered the initial paradigm of the Web 1.0 drastically. Different digital monopolies managed to capture 95% of the overall usage with only 20% of the available APIs at the given time in 2007 [SO-13]. This superior market position of a few internet giants resulted into the alteration of APIs to direct data primarily into monopolies. This does not allow in a terms-of-service-complaint way to easily thinker with valuable data that these monopolies possess [SO-21]. However, platform owners like SAP with over 13.000

partners are starting to acknowledge these concerns for their platforms: “reaching our full potential depends on how well we enable our partners” [SO-22].

Ultimately, strict boundaries of any kind introduced by ecosystems might not be feasible to further enable innovation across different branches or ecosystems. As research on digital platform shows, innovation is getting bottlenecked through platforms by enforcing new boundaries and therefor limit interaction and growth of ecosystems [SO-22]. Alternative decentralized approaches through defining new standards like the signal messaging protocol [SO-26] or the mastodon protocol [SO-27] are ongoing efforts to bring back the Web 1.0 paradigm with recent technological advancements in security and interoperability.

Finally, end users are dependent on utilizing APIs needed for their use-cases. Also, governments are continually shifting their product development strategies towards external ecosystem-independency [SO-24]. In this niche, control algorithms and low-code tools start to develop monopoly-independent and decentralized ecosystems of their own [SO-23][SO-25]. With technology and branch independent architectures, behavior trees strive to host APIs for all kinds of ecosystems without the need to adapt the core architecture. Further low-code tools also try to approach a similar development but either lack the branch independency or are too simplistically designed for programmatic extension of handling different kinds of technologies or ecosystems [SO-25].

3.1.3.3. Machine Learning Operations

A specific type of services consists of those that are based on machine learning, often named artificial intelligence (AI) services. Machine learning based services take data and map it to an output however neither the developer nor the user knows the exact logic and its exact decision boundary since it was not defined by humans but defined or learnt by a statistical algorithm. Due to this fact, the operations of machine learning services, the so-called machine learning operations, raises specific needs and challenges. [SO-28]

In specific, those challenges are high for industrial-grade AI, which comes with high requirements, for example data privacy, low latency, high frequency data, high availability or trustworthy, reliable, and explainable behavior of the model.

To address these challenges, it essential to monitor AI services. Thereby, two major classes of data can be identified. On the one side, infrastructural metrics, like memory utilization, health status, response times, or latency, that can also be seen as common metrics for all types of services. However, on the other side there are also the streams of input data. Over the time, the distribution within the data may change, for example due to changes on the production, changing environmental condition or aging effects of the sensors. When monitoring those data streams, it is possible to take countermeasures like retraining the machine learning model, discussing the changes with the domain expert, or taking the service offline to avoid harmful consequences. [SO-29] Even though there are attempts to define architectures for machine learning operations for industrial purposes [SO-30] enabling this monitoring, there is until now no broad consensus.

For some industrial use cases this monitoring of input data is in specific hard, since monitoring the performance requires a label however labeling is extremely expensive, destructive, or contradicts with the business case of the AI service. Before the training, the methods of active learning are already tested to reduce the number of data points required for training a performant model [SO-30] but the application in the productive phase of the model remains so far untested.

3.1.4. Data orchestration approaches

During the Big Data era enterprises have seen a proliferation of data sources of different nature. Traditionally, data managers kept a strict control of the data available within the enterprise by defining strict data schemas, formal vocabularies, and metadata catalogues. However, the need for quickly processing and consuming high volumes of data led to movements like NoSQL that sought to avoid the rigid control of enterprise data management. The NoSQL paradigm resulted into new types of “schemaless” stores such as document databases or data lakes, which enabled agile application development as well as an easy way for data scientists to access data in the so-called “schema on read” fashion. As a result, data was mostly enclosed in the applications, thus, unaware to the rest of the enterprise. This lack of data governance had a great impact on matters like data management regulations or applications that would require a holistic view of what the enterprise knows.

Therefore, enterprises need to find a balance between understanding all the available data while promoting agile development for their applications.

To tackle these challenges, data infrastructures are now shifting towards dynamic, distributed approaches. Especially, two types of architectures are gaining traction within the industry: the data fabric and the data mesh.

3.1.4.1. Data fabric

Gartner analyst defines the data fabric as “as a design concept that serves as an integrated layer (fabric) of data and connecting processes. A data fabric utilizes continuous analytics over existing, discoverable and inferred metadata assets to support the design, deployment and utilization of integrated and reusable data across all environments, including hybrid and multi-cloud platforms” [DOA-1].

The data fabric introduces a new architectural approach that facilitates the integration of data regardless of their source location. The integration layer abstracts data consumers from the heterogeneity of the underlying stores by providing a unified view of the data. In this sense, data access to this unified view is facilitated through a centralized interface. As a result, the data fabric propels the transition to a data-centric mindset, where consumers only need to think about what data they want rather than bothering about gathering and connecting data from different sources.

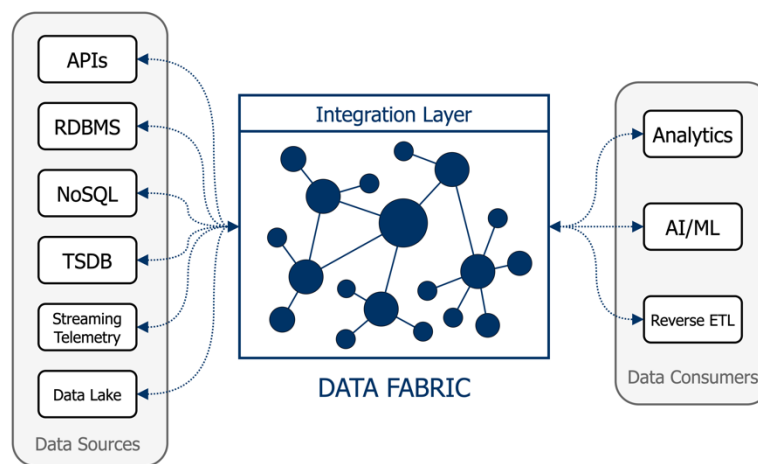


Figure 18. Architecture of the data fabric

Sitting at the core of the data fabric is the knowledge graph, which can be defined as the graph representation of the combination of business data and an explicit representation of knowledge. The term became popular after Google unveiled their vision of a knowledge graph back in 2012 [DOA-2]. Google proposed a graph of real-world entities and their relationships to one another, hence, the idea of “things, not strings”.

The knowledge graph is considered a technology that can meet the requirements of a successful data fabric [DOA-3][DOA-4]. Graph data representation enables handling connections between data among distributed sources. The explicit representation of knowledge in the graph provides semantic metadata that enables making sense of data itself. Lastly, graph representations are particularly suited for data integration as they can accommodate any kind of data depending on the nature of the source (e.g., RDBMS, document databases, graph databases).

In addition to the knowledge graph, another fundamental technology implemented by the data fabric is known as data virtualization [DOA-5]. This technology allows for keeping data at the source, thus, avoiding copying data into the data fabric for integration. The data fabric makes this possible by establishing smart indexes to the data, so that when consumers of the fabric request some data in particular, the fabric goes to the respective source system, collects the data, and returns it back to the consumers.

However, data virtualization must be seen as a double-edged sword. Virtualizing data brings many benefits such as fast integration of new source systems, saving costs in terms of storage, and facilitating data governance. On the downside, query performance could be deteriorated as every time a consumer requests for virtualized data, the data fabric federates the query to the target source system, therefore, adding extra delay in the transaction.

This limitation could be addressed by using caching mechanisms, but still, the data fabric should be flexible enough to specify which data must be virtualized and which must be materialized.

As of this writing, we found no trace of the concept of data fabric neither in the literature nor in SDOs, but it seems to be mostly a commercial term. A market analysis shows that the concept is drawing much attention in the industry as we already see several data management vendors offering the data fabric as part of their portfolio. Some of the main vendors are IBM [DOA-6], K2View [DOA-7], Informatica [DOA-8], data.world [DOA-9], Stardog [DOA-10], or Talend [DOA-11].

3.1.4.2. Data mesh

Zhamak Dehghani defines data mesh as “a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments—within or across organizations” [DOA-12].

Data mesh focuses on analytical data, so before going through the undercurrents of the data mesh paradigm, we must first understand which are the differences between operational data and analytical data.

Operational data can be defined as the data that keeps the current state of the business. Sometimes referred to as “data on the inside”, it represents the data that business applications use to serve the end users. This data is typically stored in systems like OLTP which are optimized for data access.

Analytical data is the historical, aggregated view of operational data. This kind of data is usually stored in systems like data warehouses (OLTP) or data lakes. In this case, known as “data on the outside”, analytical data is consumed by data analysts and data scientists to derive retrospective or future insights on the business. Thus, analytical data helps to optimize the business and user experience by using techniques like machine learning or business reports.

The data mesh is classified as a sociotechnical paradigm that introduces not only a new data architecture, but also an organizational operating model for the people that interact with data. This new paradigm builds upon four principles:

- **Domain Ownership:** Ownership of data is decentralized by defining business domains. This principle aims for an architectural and organizational alignment among business, technology, and analytical data. By logically decomposing data in separate domains, centralized bottlenecks like data warehouses and data lakes are removed, thus enabling scalable data sharing that can keep up with the increasing diversity of data sources, data consumers, and data use cases.
- **Data as a Product:** Domain-oriented data is managed as an asset to be shared with data users, as opposed to the traditional approach of collecting data in silos. For data to become a product it must be discoverable, addressable, understandable, trustworthy, truthful, accessible, interoperable, self-contained, and secure. Overall, data product must be autonomous, where the lifecycle management of data and its models must be independent from other products.
- **Self-serve Data Platform:** This new generation of data platform empowers cross-functional domain teams to share data. The platform implements the capabilities to manage the lifecycle of data products. The data platform enables data users to seamlessly discover, access, and use a mesh of interconnected data products.
- **Federated Computation Governance:** Data governance operating model with a team composed of domain representatives. This federated model aims to define cross-cutting governance policies across a mesh of distributed data products.

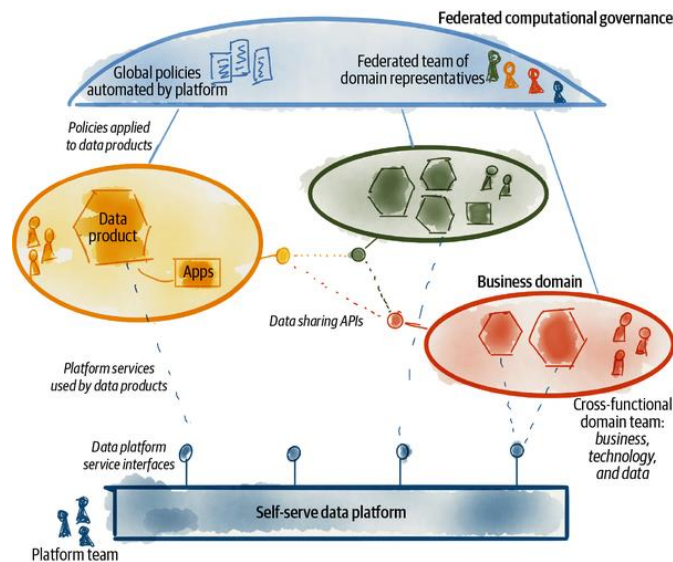


Figure 19. Simplified operating model of data mesh [DOA-12]

Data mesh and data fabric architectures are considered as complementary, being the data fabric an enabler of the data mesh [DOA-13]. The data fabric provides data owners with mechanisms for automating the creation of data products and for managing the data product’s lifecycle. The unified view of data and metadata delivered by the data fabric enables cataloguing data assets, turning these assets into products, and applying federated governance policies. The resulting data products are exposed through the unified interface of the data fabric, which data consumers leverage to search and access data products.

Even though the paradigm of data mesh is at an early stage, we find in the industry data management vendors like K2View or data.world that already commercialize implementations of data platforms aligned with the data mesh architecture.

3.2. Review of relevant techniques for the meta operating system

This section provides a review on key technological domains, and their recent advancement, which set the basis for a newly introduced heterogeneous mix of enabling technologies in the communications’ and compute domains and to their respective control, facilitating thus radical new use cases that extend from the data center core to the network edges. Rapidly growing bandwidth and low latency, cost, security/privacy requirements along with the surge in data volume that is anticipated from the massive number of devices deployed over a variety of networks and environments, are pushing for a migration from traditional cloud based data processing and computing towards an edged-based provision of services closer to the end devices and users. A seamless synergy between edge-based and cloud based services that will provide users with a use case agnostic services’ environment is dictating for a network and compute continuum. Recent innovations and techniques supporting automation and optimization across communication -and also newly introduced- computing and data service planes, addressing complex project needs are discussed below.

Within aerOS architecture **containerization and virtualization techniques** have a prominent role and significance as they are key players which enable a collection of resources (e.g., network, network nodes, storage and processing hardware) to appear to end users as a single coherent system which may seamlessly devote the required, per use case, “virtual” resources. Overlay networks, created over multiple physical connections, meet diverse needs with distinct individualized policies. Computing resources hardware and software dependencies are separated and virtualized network functions (VNF) completely transparent to the actual hardware below are implemented on top of Virtualized Infrastructure Managers (VIM) which manage underlying infrastructure components, including storage, computational and network resources. Containerization of applications enable the deployment of coarse-grained set of microservices. All these are transparent to the end user who just perceives a compute and network continuum as single coherent system. Moreover with IoT expansion more data is produced and consumed at the edge and more compute tasks are executed at the edge than ever before,

so techniques enabling migration and placement of processing tasks towards the edge devices, where data is generated, are required. **Edge native – cloud native techniques across network continuum** which will be able to support emerging use cases with extreme service requirements in a variety of sectors are discussed in the following sections. Techniques that will enable service provision to come closer to end user devices and that will also enable restricted devices be part of service execution are explored with the goal to advance the level of edge native techniques which at the moment stay far behind relative cloud native techniques. One of the main characteristics of the nodes that will be integrated in the aerOS IoT network continuum is the heterogeneity; it is this fact that makes IEs overall configuration and management & orchestration more difficult. In order to provide nodes' autonomy and independence **self* capabilities of heterogeneous nodes** are required. A list of self* capabilities is investigated and the more crucial ones for the project implementation are defined, the ones that will enable a coherent and autonomous nodes management despite their differences regarding their architecture and software.

The aforementioned, aerOS, computing continuum will integrate a significant number of data producers. For a smooth operation it is necessary that all the nodes that are going to be part of it have the capability to exchange information that can be mutual perceived and interpreted. Service based architecture will provide the possibility for data exchange and so **data interoperability** at least in a **syntactic and semantic** layer is important so that a mutual “agreement” will enforce data to follow some predefined formats and schema with a recognizable and interpretable meaning.

Also what is important, from a data perspective, is to be able to maximize their value while protecting them and preserving their privacy. Under this aspect **data sovereignty governance and lineage policies** are explored in order to ensure their quality, integrity, security and usability. Additionally, data produced across the platform will give the opportunity for **advanced AI management approaches** are explored for deployment within the cloud-edge continuum, both as services implemented for edge IoT devices and processes but also in a platform functional level, providing critical decisions for the best adaptation and needed reconfiguration of networking services. Based on the offered edge nodes capabilities, AI models which can perform on a federated basis and which make use of limited data sets are described so that, beyond the demanding powerful training operations conveyed on powerful cloud machines, training and prediction activities can be placed at the near-edge devices and users targeting better response times for critical events and less bandwidth usage.

The extended range and the multitude of services to be deployed across the cloud-edge continuum erase the need for a strong security and privacy framework. Development cycles require the extension of DevOps methodologies to a **DevSecPrivOps** and the way in which, by design, security and privacy will be included, so as to ensure agile long-term evolution, is also discussed. Moreover, beyond DevOps, a strong **decentralized security and privacy** system which will support a cross-layer cybersecurity solution for all resources management is required both due to the multitude of services deployed across the cloud-edge continuum and also due the distributed nature of data production and storage. Access to network resources, services subscriptions and utilization, data access and ownership should be governed under a holistic mutual trust assurance “umbrella”. Integration of authN/Z, logging services and interoperable control of aspects such as data usage, consent, ownership, H/W & S/W access are analyzed below.

In the next sections all above points and techniques are extensively discussed.

3.2.1. Real-time containers in the Industry

When discussing about IoT edge-cloud continuum architectures in the Industry, virtualization technologies such as hypervisors and containers take a very central role. Although these technologies are very prominently used in the cloud, they still must be adapted to IoT devices and the edge. In particular, if they are to be used in safety-related and real-time (RT) systems there are multiple research questions to be addressed and solved. Cinque et al. [RTC-1] displays the first prototypical implementation of an architecture for hosting RT-containers which guarantee temporal and fault isolation as a minimum, such that a fault of a non-critical component does not affect a critical component. Instead of a kernel patched with Preempt-RT, they proposed a dual kernel system, where one kernel is responsible for hosting hard RT applications. In this prototype containers were scheduled using Earliest Deadline First (EDF) and constant-bandwidth server (CBS) implemented in Real-Time Application Interface (RTAI). Results were produced for this scheduling scheme and compared to containers scheduled with the standard policy of Linux and showed that their architecture outperformed this standard

policy. T. Cucinotta et al. [RTC-2] contributed to network function virtualization (NFV) where their aim was to deploy virtual network functions, e.g., switches, routers etc., into real-time containers. They did not aim for hard-RT guarantees, rather they used a probabilistic analysis method better suited to the cloud context, in which NFV will find its greatest applicability. In their work, they extended the open-source cloud software OpenStack, such that it is possible to deploy containers to a Linux system. These containers were scheduled using hierarchical scheduling based on the deadline scheduler extended for containers. The containers were allowed to run on multiple cores. As communication protocol between containers, they use TTEthernet (Time-Triggered Ethernet) so they can assume constant transmission latencies and throughput. To analyze the system, they modelled not only the scheduling of containers but also communication between them including the queuing that happens in standard Ethernet networks. Furthermore, they provided for high traffic loads in their model of the system. This model is then able to provide soft-RT guarantees, since is a probabilistic model. [RTC-3] Fiori et al. propose changes to the orchestrator kubernetes that enable it to also deploy rt-containers. Their architecture provides for the use of a reservation-based scheduler, and can thus be applied to HCBS, the standard SCHED_DEADLINE policy or in the case of using a bare-metal hypervisor, like Xen, RTDS. In their tests they use HCBS into which their rt version of kubelet interfaces to set period and runtime for each cgroup. They also further adapted the interface of HCBS such that cores could be specified with 0 runtime, resulting in them not being available for executing RT-tasks. Besides containers and hypervisors, Unikernels [RTC-4] are another novel virtualization technology which offer better performance concerning runtime overhead and boot times and have a small code base and memory footprint. Unikernels don't need a separate operating system and an executable image executes natively on a hypervisor. Such image contains application code, as well as all the operating system functions required by that application.

3.2.2. Edge-native approaches: cloud-native techniques applied along the computing continuum

The technological leading companies have been focused on the development of the cloud during the last ten to fifteen years, where computing capabilities and infrastructure are offered in a central location by demand. This involves different concepts depending on the service that is requested: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Functions as a Service (FaaS). In addition, this hardware located at the cloud counts with a uniform configuration that has been completely customised by vendors in order to achieve a better global performance (e.g., the machines located in a datacentre runs the same Linux based OS that has been developed by the owner company, specifically adapted for running the products that they offer; or deploying racks of machines particularly configured to offer only one type of services like K8s clusters) and fully coupled with strong and reliable network connections. On the other hand, the devices located at the edge tier of the computing continuum are entirely diverse, not fully controlled by its owners (some of them are vendor locked and cannot be customized) and are deployed in a field where the network connections are often not reliable. For that reason, nowadays cloud computing has a grade of maturity that has not yet been achieved at the edge. It is fair to state that the stage of the edge computing is, now, at the same degree where cloud computing was 10 years ago. Some cloud native techniques (e.g., virtualization, containerization, container orchestration, microservices, DevOps and CI/CD pipelines), which are vendor-agnostic de-facto standards (despite there exist particular implementations per provider - e.g., cloud providers have its own Kubernetes distribution but all of them accomplish with the standard reference and there are solutions to use clusters of different providers), do not have their equivalent edge native techniques or are in an early development stage. The challenge, then, is to create edge-native techniques relying on vendor-agnostic de-facto standards in the same way that it has been created for the cloud, and as much as possible, to recycle or adapt cloud-native ones to the edge, taking advantage of their long-term production stage.

3.2.2.1. Usage of containers in the edge

One of the main differences between edge and cloud is that in the latter the hardware resources are nearly unlimited because a large number of powerful servers located on the well-known datacentres are connected in order to achieve it. At the edge, a more heterogeneous range of devices live, which usually are resource constrained and equip a wide range of CPU architectures. These hardware limitations make it difficult to run the Docker Engine for some devices, which is the most used container runtime and the de facto standard for the

container virtualization. Nevertheless, there have been developed some solution for running containers in embedded devices with resource limitations.

The container engines are trying to reduce their memory footprint through the reduction of their internal components and the removing of no necessary modules in order to increase their performance which appears to be an interesting trend for both cloud and edge. This has become a tendency since Docker released its low-level runtime, runc [ENA-1], as open source, because the low-level container runtime is only a part of a container engine, it is the block that finally runs the container inside the system and interacts with the high-level container engine. In the same way than Pantavisor, crun is a low-level container runtime fully compliant with the Open Container Initiative (the organization in charge of the containerization standards) runtime specification that is written in C [ENA-2], in contrast with runc, which is written in Go. Furthermore, the crun binary is smaller than the runc binary (300KB versus 15MB) and has proved a better performance in a test consisting of running sequentially 100 containers, where crun achieved the test goal in 1,69 seconds while runc took 3,34 seconds [ENA-2]. Focusing on the container engines, an interesting container one could be cri-o, a lightweight container engine specifically designed for Kubernetes [ENA-3]. Its main advantage is the reduction of the resources consumption in comparison with Docker, Moby or rkt. However, cri-o has not been developed for its usage in a low-resource embedded device, the target device of cri-o must be able to run Kubernetes.

BalenaOS is a host operating system with the purpose of running Docker containers on embedded devices, specifically fine-tuned for containers and made to survive harsh networking conditions and unexpected shutdowns [ENA-4]. This OS is based on Yocto Linux, which provides a small memory footprint and the possibility of an easy porting to more powerful devices across a set of varied CPU architectures. Its purpose is to achieve the development of applications in the same way that if these applications would be deployed in a cloud operating system through the usage of containers. For this purpose, Balena has developed its own container engine (balenaEngine) based on the Moby project technology delivered by Docker and through the removal of heavy Docker features more oriented to the cloud (Swarm, plugin support, overlay networking drivers...) that are not really needed for embedded devices of the IoT world and also through the addition of specific features for this kind of devices. At this moment, the OS is supported by up to 20 device types.

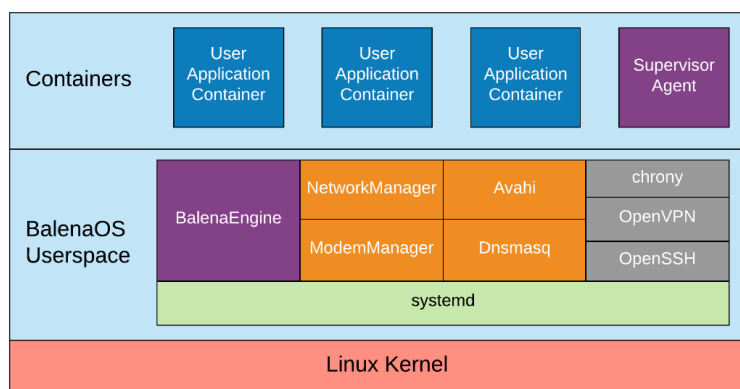


Figure 20. BalenaOS block architecture [ENA-4]

The Balena company offers an automated platform hosted by the company itself to manage the infrastructure running BalenaOS and the workloads deployed in such devices that is named Balena Cloud, which has been optimized for the edge. Last but not least, Balena has delivered this management software in an opensource way for advanced users or infrastructure managers that want to host this platform in their own infrastructure without depending on Balena. By using this platform, developers are capable to deploy application containers, push updates, check status and view logs of the fleet of devices that has been previously registered.

According to the enterprise Pantacor, Docker has not been built having in mind the embedded devices due to its high resource requirements, for that reason, they have developed Pantavisor, a minimal low-level container runtime written using the C programming language, like the LXC containers (LinuX Containers were the first developed container-based virtualization method) and the Linux Kernel [ENA-5]. Pantavisor has the purpose of evolving the traditional embedded systems, from the legacy monolithic firmware and applications to a modular approach using pure Linux containers, or in other words, into a set of portable and reusable

microservices. According to Pantacor, Pantavisor “is meant to be a single-binary init system that boots directly from the kernel and becomes the first process to run, which then brings up the rest of the system as a set of well-defined micro-containers” [ENA-6]. This container runtime is compatible with Arm, Mips, Risc-v, x86 and PowerPC CPU architectures, its average size once running on a device is only of around 350 KB and requires a minimum of 64 MB of RAM.

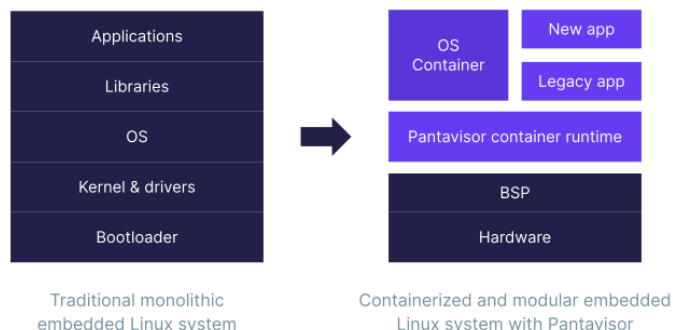


Figure 21. Evolution of embedded systems from the point of view of Pantavisor [ENA-5]

In contrast with the traditional container-based architectures, Pantavisor doesn’t need a complete OS running on top of a container runtime. This is extremely beneficial to the embedded devices because they are not really using all the features of a host OS, so Pantavisor containerizes the host OS layer, which becomes a container with the same characteristics of an application container with the advantage of having the ability to be updated in a straightforward way, and finally Pantavisor acts as the minimal container runtime manager of the system. In the same way than Balena, Pantacor provides a framework to manage the devices running Pantavisor and the workloads deployed inside them that is named PantacorHub, which is both offered opensource for self-hosting and as a paid service hosted by the company itself.

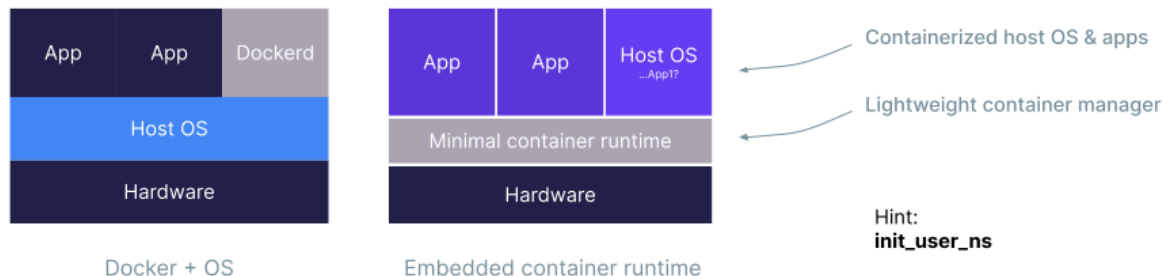


Figure 22. Architecture comparison between traditional container-based and Pantavisor-based [ENA-6]

Finally, EVE-OS is an Operating System originally developed by ZEDEDATA and then donated under an opensource license to the Linux Foundation that has included it inside its Edge researching projects stack [ENA-7]. The main purpose under the development of this OS is to provide to the edge computing equipment a universal, vendor agnostic and standardized architecture OS following the same strategy that used Google in the smartphone market when they delivered Android. EVE has adopted well-known opensource projects like Xen Project, Linuxkit and Alpine Linux for its development. Currently, the management of a fleet of devices running EVE-OS is possible using the Adam controller, the reference implementation of an LF-Edge API-compliant Controller. Furthermore, EVE provides support for running containerized workloads using a container engine and for Kubernetes distributions. Last but not least, its main difference with BalenaOS and Pantavisor, is that EVE-OS is designed for edge devices with more powerful equipment and not embedded systems with less than 512MB of RAM but allows its deployment in a wide range of CPU architectures. Moreover, EVE has enough capabilities to be deployed on bare metal and supports a wide range of workloads that can be combined: Docker containers, Kubernetes clusters and virtual machines.

3.2.2.2. Container orchestration at the edge

Kubernetes (K8s) has become the standard for container and microservices orchestration in the cloud, advantaging its competitors in the last years like Docker Swarm or Apache Mesos. In addition, the vast majority of the public cloud providers has delivered its own Kubernetes distribution that are fully compliant with the K8s standard references in order to optimize the integration of K8s in their systems. One of the trends in the last years has been to take containers to the edge computing deployments, so if Kubernetes is the standard for the container orchestration, it must be deployed at the edge, at least in the intermeddle nodes of the edge or, in other words, in devices that are resource constrained but have enough capacity to carry out some more powerful workloads in comparison with the leaf devices of the far edge tier. To achieve it, there has been appeared some Kubernetes distributions or K8s based solutions optimized for the edge.

K3s is a lightweight fully compliant Kubernetes distribution developed by Rancher focused for running in constrained devices, as its memory footprint is much lower than other available K8s distributions [ENA-8]. K3s modifies the K8s paradigm of master and worker nodes, converting them into server and agent nodes. In addition, offers three possible architectures: a single server with an embedded SQLite database and high-availability servers using an embedded or an external database (SQL based database or etcd). This distribution is as well optimised for ARM32, ARM64 and ARMv7 platforms, hence better in case of leveraging common embedded systems as nodes like Raspberry Pis or NVIDIA Jetson boards. Its minimum requirements are 256 MB RAM usage for an agent node and 512MB for a server node with some workloads running in an agent node. Rancher have also delivered an operating system optimized for running K3s with only the minimal resources of the underlying OS: k3OS [ENA-9]. This low memory footprint and its ability to run in devices having diverse CPU architectures converts K3s into the most recommended K8s distribution for building clusters at the edge. On the other hand, Canonical has released MicroK8s, another lightweight K8s distribution with a minimal memory usage of around 540 MB, but its recommended memory allocation is 4GB which is still notably higher than K3s [ENA-10]. From our experience, Microk8s has been tested in environments with only 1GB of memory available and this K8s distribution doesn't work properly in these constrained devices, even it can't boot up in some cases. However, MicroK8s has the advantage of an easy customization through the installation of external addons with only executing its "install" command. The available addons includes some K8s widely used modules like CoreDNS, Helm or Istio and the possibility of achieving K8s High Availability in an easy way. These features make MicroK8s one of the most interesting K8s distributions for development and testing in slightly more powerful edge devices.

Another trend in bringing Kubernetes closer to the edge tier of the edge-to-cloud continuum is to adapt K8s to the specific characteristics of the edge (unstable network connectivity, difficulty of managing heterogeneous and low-resource equipment), maintaining all its benefits achieved on the cloud and not only creating another new K8s lightweight distribution. Some of these solutions try to maintain the control plane of the system at the cloud and move only the needed workloads to the edge, converting it in an autonomous component of the system regarding the application plane.

KubeEdge is an opensource framework built on top of Kubernetes with the main purpose of bringing the full functionalities of Kubernetes to the edge that is under the umbrella of the Cloud Native Computing Foundation (CNCF) [ENA-11] [ENA-12]. The main idea under this technology is to move all the control plane to the cloud, where the computing resources are higher, and leave to the edge the workloads or the application plane in order to dedicate all the constrained computing resources of this tier for this purpose and as well as controlling the communications with the far edge devices without real computing capabilities (sensors, cameras, ...). This is translated into a low memory footprint of the EdgeCore installation of only 70MB. The KubeEdge architecture is divided into three layers:

- Cloud: at the cloud tier is needed a running K8s distribution that interacts with the also deployed in this tier CloudCore, which includes controllers to synchronize the status of all the edge nodes and the devices connected to the nodes.
- Edge: the components deployed inside the EdgeCore handle communication between application containers, connected devices and the cloud tier. The K8s pods are deployed in this layer, but its deployment is controlled by the cloud. The principal novelty is that it's not a K8s node and does not include the K8s API or its control plane. Moreover, a MQTT broker is needed to interact with the device mappers (the available mapper types are Bluetooth, Modbus and Opcua, but a Go library is provided to

allow developers to create mappers for other protocols) that are in charge of the interaction and control of the leaf devices, as well as of its lifecycle management.

- **Devices:** leaf devices with almost not computing capabilities. They interact with the edge layer using different industrial protocols for data exchange.

Moreover, KubeEdge can lead with poor network connection between cloud and edge and run the needed synchronizations only under conditions of network stability. This is a key issue for edge native applications that is not resolved in K8s because it is mainly focused on the cloud. In addition, KubeEdge also provides service mesh capabilities for the services deployed in all the edge layers controlled by the same cloud.

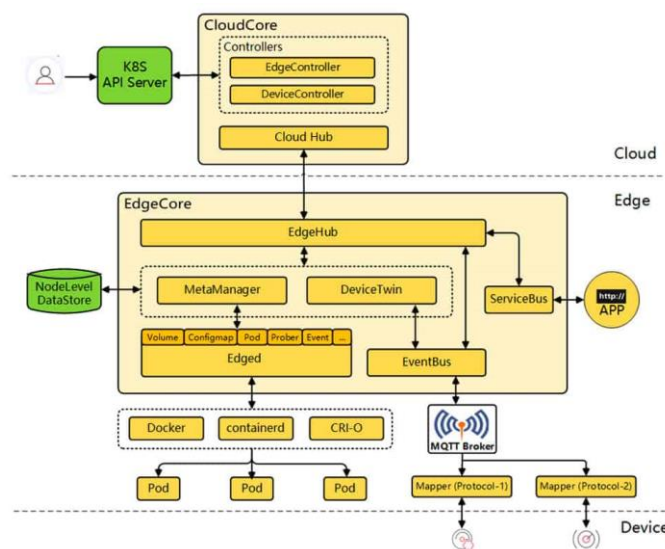


Figure 23. KubeEdge architecture [ENA-11]

An illustrative use case built using KubeEdge is the deployment of a large number of monitoring devices across the Hong Kong–Zhuhai–Macao bridge. The edge tier of KubeEdge is deployed in every device and all these devices are managed in a centralized way by the cloud part deployed in a public cloud datacentre. Each monitoring device gathers data from 14 different sensors (CO₂, PM_{2.5}, temperature, humidity, ...) through its specific mapper and the data is processed locally using AI inference programs deployed on the edge nodes (K8s pods running inside each device). Only the selected data is finally uploaded to the cloud through a reliable 5G connection, but in case of network issues there has been added a cache strategy at the edge for assuring that no data is lost during the process.

Taking advantage of the edge to cloud synergy achieved in the KubeEdge project, the same community of developers tried to use this technology to improve the execution of Artificial Intelligence workloads through the edge-to-cloud continuum. For that reason, they have developed Sedna, a project focused on implementing across edge-cloud collaborative training and collaborative inference capabilities [ENA-13].

Another interesting technology for bringing container orchestration to the edge is the one that first was developed by IBM and then donated to the Linux Foundation: Open Horizon (OH) [ENA-14]. This technology shares with KubeEdge the concept of moving the workloads to the edge tier of the architecture but maintaining the control plane (application and edge devices management) in the cloud or in a centralized environment. Furthermore, OH promises the support of the management of up to 10.000 edge devices simultaneously from a unique Management Hub instance. Open Horizon architecture is divided into two main components:

- **Management Hub:** located at the centralized cloud in which must be running K8s distribution, it oversees the control plane regarding the deployments and the edge nodes and devices.
- **Edge Agent:** this component is divided into two subtypes depending on the workload type that will be running in the node. The Edge Device Agent is targeted for resource constrained devices which are capable to run containerised workloads through a container runtime while the Edge Cluster Agent is appropriate for equipment in which can be installed a K8s distribution.

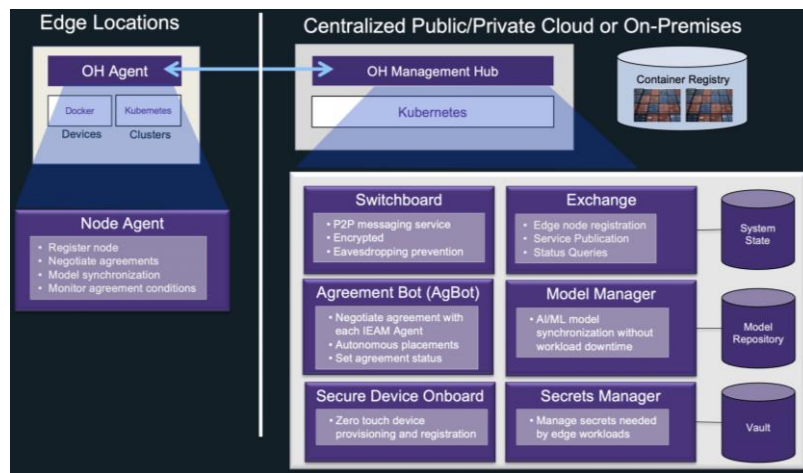


Figure 24. Main Open Horizon components [ENA-14]

Related with the EVE-OS introduced in the last subsection, project EVE’s developers are planning to support Open Horizon based workloads in the same way that K8s ones are natively supported, this is due to both projects shares belonging to the Linux Foundation.

Baetyl is another project which shares some key concepts with KubeEdge and Open Horizon since its architecture is split into the Cloud Management Suite and the Edge Computing Framework [ENA-15]. However, Baetyl only supports edge nodes with a minimum of 1GB of RAM that are capable to run a K8s distribution (K3s is recommended for resource constrained environments), it does not support the single container mode without K8s. This tool is also included inside the stage 1 of Linux Foundation Edge, so its development is in a preliminary stage with a clear lack of documentation.

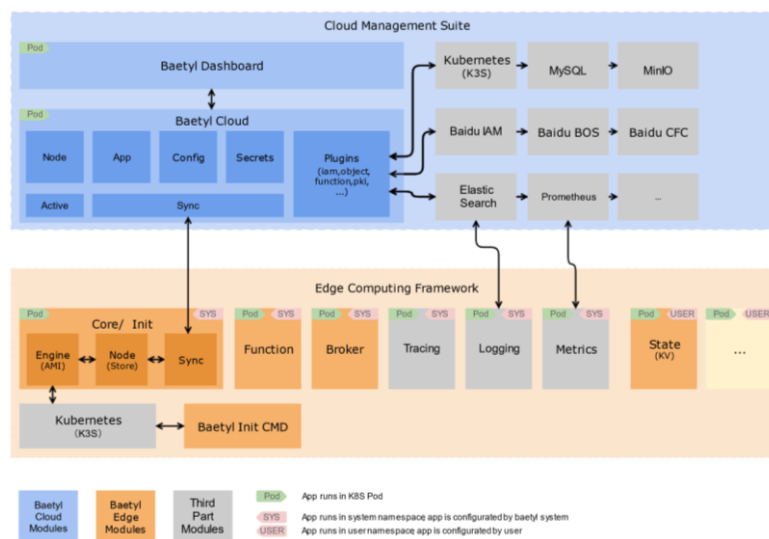


Figure 25. Baetyl architecture [ENA-15]

The Akri project that is under the CNCF umbrella as a Sandbox project, has the purpose of developing a Kubernetes Resource Interface that allows to expose the heterogeneous range of leaf devices located at the lowest tier of the continuum as resources in a K8s cluster like IP cameras or USB devices connected to the same machine that is running a K8s node [ENA-16]. This is the main difference between it and KubeEdge, Akri is a complement for K8s (provides a layer of abstraction for the devices in similar way the CNI does for the network) where the devices interact with the Akri Agent service running on the nearest K8s node of a cluster, so Akri extends the K8s functionalities but does not adapt it to edge native scenarios in the same way that KubeEdge, which tries to adapt or rebuild K8s to put it closer to the edge requirements. Following with the description of how Akri works, this technology supplies a set of device Discovery Handlers based on ONVIF, udev, and OPC

UA as well the possibility of extending this set with custom handlers. When a new device is discovered by the handlers, Akri creates a K8s service to monitor its state and gives the capability to provide high availability in the case that a node loses network connection or is broken.

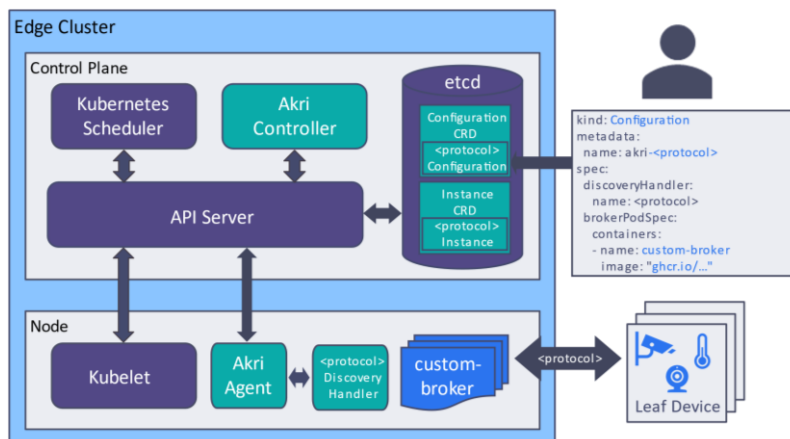


Figure 26. Baetyl architecture [ENA-15]

3.2.2.3. Serverless at the edge

Serverless architectures are widely adopted in the public cloud, offering to the customers the possibility to run their developed applications without having care of the infrastructure where it will be really deployed and only during a specified timeframe. The main advantage of using serverless at the edge is the possibility of running only the functionalities that are required in each moment, by creating functions on demand and scaling to zero when these functions are not needed. This is translated into less resource and power consumption, specifically indicated for the resource constrained devices which are present at the edge tier of the computing continuum.

The serverless paradigm has arrived at Kubernetes through projects like OpenFaaS [ENA-17] and Knative [ENA-18], which is an incubating project of the CNCF and the newest and the most interesting technology inside this scope. Knative also provides a complete event driven engine based on CloudEvents [ENA-19], a specification to standardize event data descriptions, which opens a wide range of possibilities in K8s based architectures and obviously to the edge. With the inclusion of this event driven engine, the deployed microservices can throw different events to activate some functionalities or workloads without the modification of its source code.

3.2.2.4. Public cloud providers approaches

The vast majority of cloud providers have treated its edge computing solutions like an extension of its own cloud infrastructure and commercial solutions but located at the customers premise. In addition, these providers have constructed its own hardware devices that are completely vendor locked to constrain their working scope to their own cloud infrastructure, so it provides the great advantage of being plug-and-play devices with zero-configuration needed from the final user. These are its main advantages; however a great part of these solutions is not actually edge native solutions that follow the new edge natives approaches and requirements, they are an strategy of moving workloads outside the cloud infrastructure but without breaking a strong dependency on the cloud.

Amazon Web Services (AWS) offers a solution named AWS IoT Greengrass for deploying processing capabilities in devices across the edge tier, specially IoT devices that gathers data from different attached sensors [ENA-20]. This solution includes serverless based deployments in the devices using AWS Lambda, the serverless approach of AWS, container-based deployments and AI inference capabilities at the edge through the usage of created and programmed models at the cloud. To avoid problems related to network connections, Greengrass creates a virtual twin of the device that is constantly checking the real status of the device with the desired one, so this status is only synchronized when the network connection with the cloud is reliable. In addition, inter device commutation is allowed inside a local network without depending on the cloud. Related with device management, it has the ability to be completely configurable remotely to add and remove modules

in order to avoid memory restrictions or extend its capabilities with a great customized software catalogue. Another solution from AWS for the edge is AWS Snowball Edge, a device type from the AWS Snowball family that is designed for working at the customer installations [ENA-21]. Amazon offers three types of devices: storage optimized for data transfer (80GB of usable storage capacity), storage optimized with EC2 compute functionality (AWS product focused on proving computing capabilities on demand) and compute optimized (with an average of 104 vCPUs and 416 GB of memory). Snowball edge devices can be managed locally and through the cloud, furthermore they can run powerful workloads to move all its stored local data to AWS S3 storage service at the cloud and control the deployments running at the Greengrass devices. Finally, these devices could be classified at the top layer of the edge tier of the computing continuum, because provides great computing capabilities near to a small cloud datacentre.

Microsoft Azure provides a complete stack for the edge computing under its Azure IoT Edge framework, with remote equipment management, edge level virtualization and remote workload allocation and control [ENA-22]. This technology is delivered with an opensource MIT license at the GitHub account of Azure to allow developers to deploy and integrate all its edge stack with the customers infrastructure, nevertheless this stack is finally depending on the Azure cloud stack, that is the opposite to opensource. Azure also has certified a vast range of the lowest and the medium tier of the edge tier devices (some of them are based on popular embedded devices like Raspberry Pis, NVIDIA Jetson boards and INTEL boards) what means that experts have validated that a device “can connect with Azure IoT Hub and securely provision through the Device Provisioning Service (DPS)” [ENA-23]. What’s more, following the same strategy than the AWS Snowball edge devices, Microsoft offers a line of powerful equipment to bring all the Azure service to the customer installations and avoiding the uploading of heavy workloads to the cloud under the name of Azure Stack Edge. This equipment is divided into two lines of products: Edge Pro Series, a line focused on powerful products to be located both in a local datacentre (Pro and Pro 2) and a transportable equipment that can contain an uninterruptable power supply (Pro R); Edge Mini Series, a constrained portable device operated by a battery [ENA-24].

Google Cloud integrates edge computing solutions under its Google Distributed Cloud solution. One of the most interesting features of this solution is that Google a really edge-to-cloud continuum by offering services of all the tier of the edge-to-cloud continuum architecture layers, including the telecommunication service provider network layer and the possibility of virtualization of the telcos 5G network elements [ENA-25].

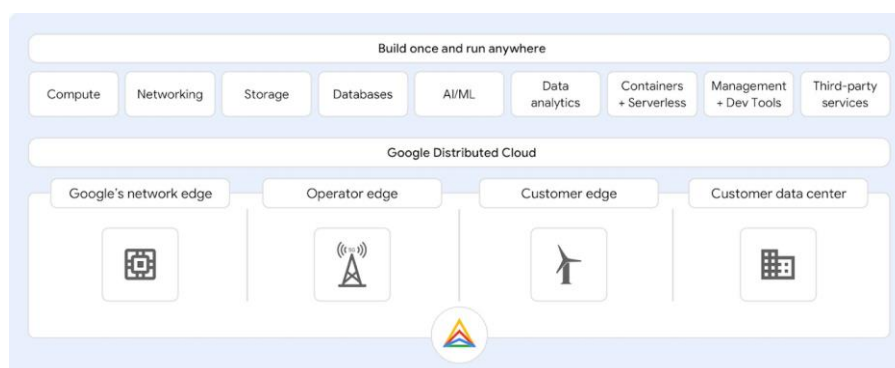


Figure 27. Google distributed cloud architecture [ENA-25]

3.2.2.5. Alternatives to containers

In the previous sections there have only been presented technologies that rely on containers for virtualization due to containers are the de facto standard for running virtualized deployments at the cloud and also are the natural evolution of the legacy virtual machine (VM) based deployments, moreover, since Docker was released in 2013, this virtualization technique has been successfully tested in deployments around industry’s public and private clouds for different purposes. Nevertheless, containers are not perfect because they present some weaknesses (e.g. in security) and a reduced capacity for improvement. For that reason, other virtualization techniques and future tendencies that could compete with containers in the short time have been appeared.

First, containers were created to replace VMs, but an interesting capability of the latter are that provide a better isolation because don’t share the kernel of the host machine (each machine has its own kernel) and are hypervisor isolated, this separation occurs at a lower level than in containers. This could solve the challenges in

securing user workloads based on containers within multi-tenant untrusted environments. For taking advantage of this, VMs are being reduced to achieve the so called microVMs or lightweight VMs that are faster and lighter as containers. Kata Containers is an OpenStack technology that makes it possible, its main goal is to run lighter VMs instead of containers using its fully OCI compliant container runtime, which means that the popular OCI image specifications like Dockerfiles are compliant with this runtime and can be run natively using the Kata Containers approach [ENA-26].

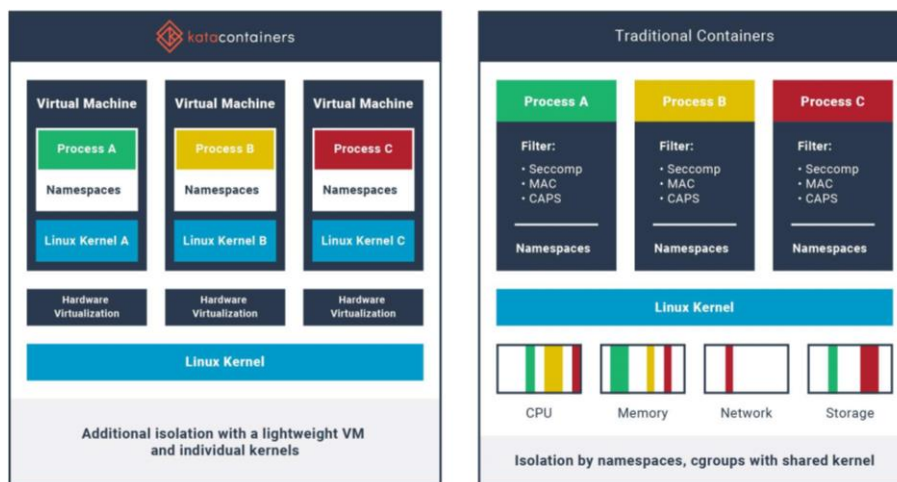


Figure 28. Differences between Kata Containers' VMs and traditional containers [ENA-26]

Another approach to replace the containers with microVMs are Unikernels. According to [ENA-27] “Unikernels are single-purpose appliances that are compile-time specialized into standalone kernels, and sealed against modification when deployed to a cloud platform”. The main idea beyond Unikernels is to only use the strictly necessary part of the user and kernel space of an operating system to obtain a customized OS that will be run by a hypervisor without the need of a host OS. This is translated into a reduction of images size, their booting time as well as their footprint and their possible attack surface. However, this virtualization technology has many disadvantages, the main one is the lack of standardization compared with containers, followed by the limitation of debugging and monitoring capabilities [ENA-28]. As an example, MirageOS is a library operating system that builds Unikernels using the OCaml language together with libraries that provide networking, storage and concurrency support [ENA-29]. Nabla containers is an IBM research project focused on building a platform to handle Unikernel workloads (for instance, workloads built using MirageOS) through the usage of its OCI compliant low-level container runtime runc [ENA-30]. Its main limitation is that Nabla is not OCI image spec compliant, so it is not able to run software that is packaged using other container image specification other than Nabla specific built ones.

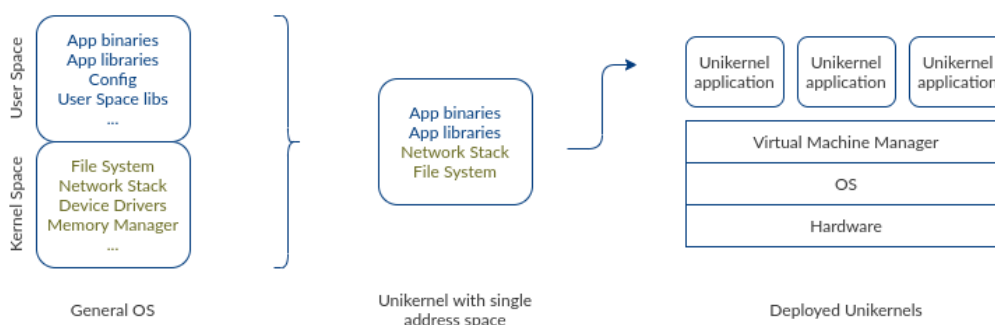


Figure 29. Creating and running a Unikernel [ENA-28]

Finally, the most promising and novel trend to be established as a strong alternative to containers is the one based on WebAssembly (Wasm). Wasm is “a binary instruction format for stack-based virtual machine” developed by World Wide Web Consortium (W3C) which allows that software written in a set of different languages (C++, Go, Kotlin, ...) can be compiled and executed with a nearly native performance in web

applications that are designed to run in the web browsers [ENA-31]. However, in the recent times developers that were aware of its main advantages started investigating if its promising capabilities could be moved outside web browsers, or in other words, to the server side. Its low memory footprint, improved security and isolation, fast booting (up to 100 times faster than containers) and response times makes Wasm perfect for running workloads in edge computing devices that are not capable to run container workloads or for resource constrained environments, where could lead to an increase of the simultaneous running workloads compared with the number that was achieved with containers. This research led to the creation of the WebAssembly System Interface (WASI), a “modular system interface for WebAssembly” with the main purpose of enabling the execution of Wasm in the server side through the creation and standardization of APIs that must be independent of the used Wasm engine. When Wasm is executed in the web browser, it uses the web APIs provided by the browsers to enable its interaction with external components. Nevertheless, when Wasm is run outside the browser, these standard set of APIs don’t exist yet, so this is the target of the WASI, the creation of a standardized set of APIs to really make Wasm portable across different platforms and its engines. Nowadays, WASI is still being standardized in a subgroup of the WebAssembly Community Group of the W3C [ENA-32].

In the present, there is available a wide set of Wasm engines that perform the execution of Wasm workloads. The most promising are: Wasmtime [ENA-33] and WasmEdge [ENA-34]. Wasm engines should be compared with container low-level runtimes (e.g., runc) because they can be managed by high-level container runtimes like containerd or cri-o, which in addition can act as the K8s CRI in order to allow the deployment of Wasm based workloads in Kubernetes in a transparent way for the user. Furthermore, Docker announced in October of 2022 the compatibility of the Docker engine with Wasm deployments through the usage of WasmEdge as the Wasm engine together with the containerd-wasm-shim that is in charge of the communication between the high-level container runtime (containerd) and the low-level runtime engine (WasmEdge) [ENA-35].

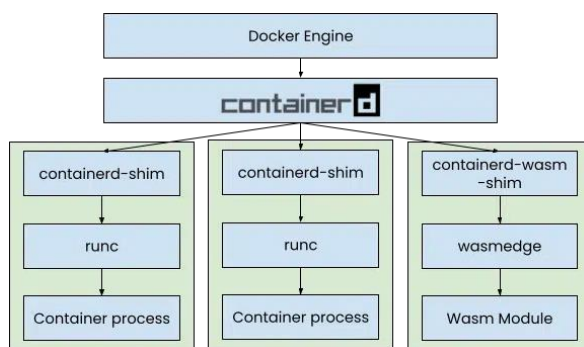


Figure 30. Architecture of the Docker Engine for running Wasm workloads [ENA-35]

3.2.3. Self-* capabilities of heterogeneous nodes

3.2.2.1. Context

Today, cloud-computing is one of the most widespread and used ways to perform complex calculations that require a large number of computing cycles, or for the analysis and processing of large amounts of data that require the highest possible speed of execution. Also, it is considered one of the most important changes in the field of information technology (IT) for society [SELF-1]. The National Institute of Standards and Technology (NIST) of the United States Department of Commerce defines cloud-computing as a model for enabling anywhere, convenient, on-demand network access to a set of shared and configurable computing resources (servers, storage, services, etc.) that can be provided and release quickly and with very little effort. This model is mainly composed by three service models: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [SELF-2]. This type of computing has many advantages over other types, however, it also has some shortcomings that are difficult to solve for certain situations.

On the one hand, the scalability, the large amount of data that it is capable of processing or the practically unlimited processing and execution capacity of calculations are some of the characteristics that make cloud-computing a valid solution for most cases [SELF-3]. On the other hand, this great work capacity requires big computing centres that are generally far from the source of data generation. This produces some disadvantages (among others) such as high latency and low response time as the information has to travel through many points

throughout the network [SELF-3]. These drawbacks prevent calculations and data processing in real time, with a very low response time and close to the source of information. Moreover, these data centres consume a lot of energy, generating a huge carbon footprint. This high energy consumption has become a big problem today because the use of energy is not efficient enough and is not always generated with renewable energies [SELF-4]. In fact, depending on the geographical area, the energy mix differs from the rest. On the one hand, in countries where emissions regulations are tougher, renewable energies predominate. On the other hand, in countries with more lax or non-existent emission regulations, fossil energy is usually the predominant one in the energy mix.

In order to carry out these operations in real time, with very low latency and greater security in the transfer of information, the edge-computing paradigm was created. This allows calculations and data processing to be performed on nodes at the edge of the network, rather than on nodes in the cloud. In this way, all the information that is produced in the edge nodes is also processed in them. This makes it possible to reduce the workload of the data centres, avoids network congestion and reduces the execution time of the time-sensitive applications [SELF-5].

There are currently several ways to define edge-computing. The Edge Computing Consortium defines it as an open, distributed platform at the edge of the network, close to data sources and integrating compute and data storage capabilities [SELF-6]. For Zhang et al. [SELF-7], edge computing is a novel form of computing that allows the storage and the processing of resources near the source of the data, providing intelligent services that collaborate with cloud-computing. Shi et al. [SELF-5] defines edge-computing as enabling technologies that allow computations to be performed at the edge of the network, at the proximity of data sources. These nodes not only consume data, they also produce and process it.

In order to create a edge-computing continuum network, it is necessary that all the nodes that are going to be part of it have the capacity to work and coordinate together. There are different types of nodes that are capable of connecting to the continuum, organized according to the network to which they belong. This classification of nodes will be developed in the next section.

On the one hand, Razzaque et al. [SELF-8] comment that one of the main characteristics of these nodes is the heterogeneity. On the other hand, Xiao et al. [SELF-9] state that this heterogeneity of the nodes makes their configuration more varied and their physical conditions more complex and changing, making their orchestration difficult. Due to this great difference in node types, each one with its own architecture and software, it is essential to have a system that is capable of executing in the same way regardless of the platform. This system not only has to be able to connect these nodes with the edge-computing continuum, it also has to be able to manage them automatically so that each and every one of them has autonomy of use.

This independence in computing nodes is achieved when the common system that governs them all is capable of offering self-* capabilities. There is a wide variety of self-* capabilities, organized and named in different ways depending on the chosen criteria. In [SELF-10], IBM explains that the essence of an autonomous system is self-management. The four main aspects of self-management are:

- Self-configuration: autonomous systems are capable of configuring themselves and their components following high-level policies.
- Self-optimization: the capacity of continually improve their performance by monitoring and identifying their resources to become more efficient.
- Self-healing: automatic diagnosis and resolution of hardware and software faults.
- Self-protection: the ability to anticipate and avoid problems and autonomously defend against external attacks or internal failures with self-healing measures.

Berns et al. [SELF-11] define a more complete list of self-* capabilities, which are: self-management, self-stabilization, self-healing, self-organization, self-protection, self-optimization, self-configuration and self-scaling. They also include two new self-* capabilities:

- Self-immunity: the system is capable of restoring security predicates after an attack, eventually preventing them from being compromised again.
- Self-containment: the ability to keep functional parts of the system not compromised by a malicious attack.

Sterritt et al. [SELF-12] define a list similar to [SELF-11] of self-* capabilities by completing it with the following: self-anticipating, self-assembling, self-awareness, self-chop, self-critical, self-defining, self-governing, self-installing, self-reflecting, self-similar, self-simulation and selfware.

For this project we have decided to use the following self-* capabilities:

- Self-awareness.
- Self-orchestrated.
- Self-diagnose.
- Self-healing.
- Self-scaling.
- Self-configuration.
- Self-optimisation.
- Self-adaptation.
- Self-learning.

The practical application of these self-* capabilities should allow autonomy of use and awareness of the environment.

3.2.2.2. Types of nodes able to be part of the continuum

The computing continuum (also called digital continuum or the transcontinuum) is the combination of resources and services at the centre of the network (cloud), at its border (edge) and in transit (fog). Data is generated and preprocessed at the edge, partially processed by intermediate nodes and, if necessary, transferred to the cloud [SELF-13]. Today there is a wide variety of nodes that are able to connect to the continuum. Each of these nodes have different characteristics and architectures that make them unique. There are several ways to classify them, depending on their architecture, type, location on the network, etc. For this project we have decided to classify the nodes according to their spot on the continuum:

- Cloud nodes: high-performance servers and high-capacity storage systems that provide services to their users. They allow complex calculations to be executed and are capable of permanently storing a large amount of data [SELF-14].
- MEC nodes: smart nodes that enable the capabilities of cloud services closer to the devices of the users. This intelligent nodes can be standard IT servers and the network devices inside or outside of the base station [SELF-15].
- Edge nodes: any device with compute, storage and network-attached capability, capable of dividing and distributing large amounts of work. Examples of these devices are access points, routers, base stations, etc. [SELF-16].
- Far-edge nodes: hardware devices capable of running algorithms that collect and preprocess information received from IoT devices or versatile computing nodes [SELF-17].
- Versatile computing nodes: geographically distributed physical devices closer to the end user such as personal computers, laptops, smartphones, tablets, wearables, smart cards, smart vehicles, etc., with sufficient computing power to execute tasks [SELF-14].
- IoT nodes: physical devices such as sensors, readers, surveillance cameras, actuators, embedded devices, etc. They are able to detect events or characteristics of real objects and transmit them to the upper layer for processing [SELF-3][SELF-14].

3.2.2.3. Self-* capabilities

In this section, all the self-* capabilities necessary to achieve independence and autonomy of use of the system will be described.

3.2.2.3.1. Self-awareness

Göttinger et al. [SELF-18] define self-awareness as an ability of computer systems to observe and analyse the environment that surrounds them and themselves, with the aim of making changes in their behaviour according

to the observations made. They also comment that self-awareness is the base in an autonomous system for all other self-* capabilities. In [SELF-19] the authors explain that obtaining knowledge of the environment can be through the analysis of the execution time of tasks, learning or sources external to the environment. In systems with hierarchies, knowledge can be affected due to the loss of a part between higher and lower levels. Esterle and Brown [SELF-20] state that the nodes of a network must be aware of other systems and devices further away from their immediate environment.

Lewis et al. [SELF-21] describe five levels of self-awareness of networked systems:

- Networked stimulus-awareness: allows the system to know how to respond to events in its environment with the stimuli received.
- Networked interaction-awareness: determines that the stimuli received and the actions performed form relationships with the surrounding environment.
- Networked time-awareness: it obtains information about historical stimuli in order to predict future stimuli and their effect on other nodes.
- Networked goal-awareness: having knowledge of the objectives, goals, constraints and preferences of the rest of the nodes allows them to know how it affects them.
- Networked meta-self-awareness: the system is capable of determining its own level of network self-awareness and how it is exercised.

In [SELF-22] Anzanpour et al. propose a monitoring and control system for the health of hospital patients with a self-aware design. This system is based on wearable devices (with limitations such as power consumption or performance) that obtain data through sensors such as heart rate, blood oxygen, blood pressure or body temperature. This information is sent to cloud servers for their storage and processing. This system provides personalized care, self-organization, autonomy of use for remote monitoring and intelligent decision-making based on the situation for patients. Andrade and Torres [SELF-23] propose a conceptual model of cognitive security, with self-awareness as the main element. This computer system is capable of generating learning models (based on self-aware knowledge) and reasoning models (created from the defined learning models).

In 2001, IBM [SELF-24] proposed a feedback loop for autonomic control called MAPE-K. This model has five phases:

- Monitor: obtain data and information from the environment for the node self-awareness.
- Analyse: the most important information obtained in the monitoring phase is selected and studied.
- Plan: the necessary actions to achieve goals and objectives are defined and built.
- Execute: the procedures for the execution of the plans are defined.
- Knowledge: the information used in the four previous phases is stored as shared knowledge.

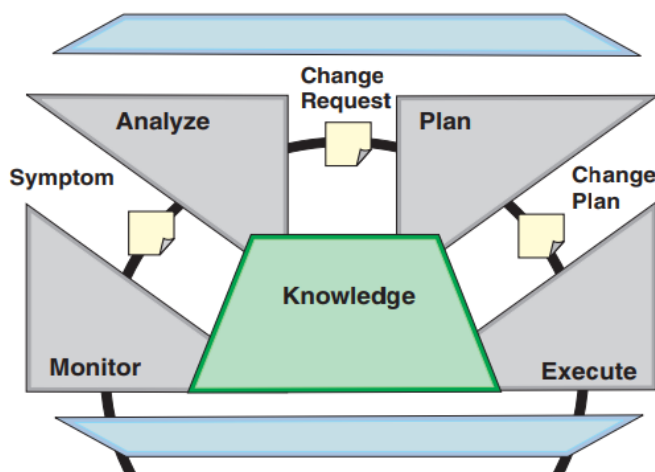


Figure 31. Phases of the feedback loop "MAPE-K" [SELF-24].

In [SELF-25], Elhabbash et al. propose a generic system that uses symbiotic simulation to address the difficulty of analysing the quality of knowledge and achieving the capacity for meta-self-awareness. [SELF-26] introduce a system for descriptive and generative dynamic models that strengthens the capacity for self-awareness. The system is based on the analysis and extension of three bio-inspired theories that have examined the capacity for self-awareness from different points of view. Zhang et al. [SELF-27], discuss cognitive digital twins, examine the concepts of digital twins and self-awareness together, and explore the possibility of harnessing different levels of self-awareness for cognitive digital twin design.

3.2.2.3.2. Self-orchestrated

Synchronous and sequential execution of services is called orchestration. Orchestration systems include the application logic needed to manage services [SELF-28]. This is one of the most important capabilities in distributed systems, because it allows applications to meet the requirements of end users. Moreover, it improves the scalability of applications and minimizes failures between the modules that make them up [SELF-29]. Based on the definition of orchestration in [SELF-30], we can define self-orchestration as the self-capability of smart devices to configure themselves, manage themselves, and coordinate with each other to achieve common goals and objectives.

In [SELF-28] Delamer and Lastra describe the difficulties in providing rapid reconfigurability in current and future manufacturing systems in the industrial sector. This is due to the introduction of new processes and devices in the systems with which already implanted components are expected to interact without having previous knowledge about the collaboration with these new devices and processes. Based on this, the authors analyse the concepts and definitions of self-orchestration and choreography oriented to web services at the node level and propose the use of self-orchestrated semantic web services to solve the problem. Khebbab et al. [SELF-30] present a rewriting-based specification developed in Maude to design and verify the self-adaptive and orchestration behaviours of the cloud and fog layers in order to manage the reconfiguration of the architecture and manage the self-adaptation and orchestration of the cloud and fog layers based on in a centralized control pattern to achieve low latency and resources quantity trade-offs.

The authors of the paper [SELF-31] propose a new reference for Building Automation Systems (BAS). This paradigm is heavily inspired by social network interrelationship models to improve self-configuration and self-orchestration of nodes in home and smart building automation. The solutions and products currently available on the market for Home and Building Automation (HBA) have limited self-configuration, automation and self-adaptation capabilities. However, these capabilities are superior to those offered just a few years ago, with very limited computing power and very slow connections. For this reason, the developed framework is based on social objects and semantic description of resources and services. This increases the autonomy of use of the devices, their capabilities to configure themselves and the relationship between them and the environment that surrounds them. These devices take on the role of intelligent agents, which can self-configure, self-coordinate, and self-orchestrate. The proposed model was implemented on Arduino boards and on Intel Edison and Zolertia single board computers with more resources.

In [SELF-32], Schulz focuses on the development of a model whose objective is to define the self-management and self-organization of a network as if it were a subsystem within automation systems. In this way, all components of the communication architecture are defined, implemented and maintained in an automated manner. The model is applied to Intranets within companies at an industrial level, orchestrating the transport of information through IP and legacy protocols as well as wired and wireless connections interchangeably. The author intends that the developed model serve as a reference for other research and as a standard in IoT networks at an industrial level.

3.2.2.3.3. Self-diagnose

Self-diagnosis is the self-capability of a smart node or device to continuously monitor its health status [SELF-33]. The node has the ability to detect the error and its origin, which allows the development of highly reliable and energy efficient applications [SELF-34]. However, the term self-diagnosis is also applied to networks made up of intelligent nodes capable of self-diagnosis or sending their health status to central nodes for further analysis. Examples of these networks can be found in [SELF-35], [SELF-36], [SELF-37] and [SELF-38].

Already in 1999, Discenzo et al. [SELF-33] evaluated the need for IoT devices for self-diagnosis of components in the industry. Thanks to a small motor together with a microprocessor, they developed a model to self-diagnose its status and prevent possible future failures. In [SELF-35], the author addresses the development of "Promising", a model capable of self-diagnosing the state of a network and its nodes. The method is based on the use of a highly reliable checking component to evaluate the state of the nodes of a network. In addition, the author recommends monitoring in a decentralized manner to minimize network traffic.

Rahem et al. [SELF-36], describe possible failures that can occur in data aggregation. This technique is commonly used to analyse and diagnose the status of Wireless Sensor Networks (WSN) due to its low power and bandwidth consumption, reduced execution time, etc. In this work, in addition, an analysis is made on the data added by the central node in the cluster to evaluate the energy consumption, using self-diagnosis. This node manages all the operations and devices that make up the group it controls. In [SELF-37], Harte et al. also develop a model to monitor the health status of nodes within a Wireless Sensor Network (WSN) using self-diagnosis. The authors focus primarily on detecting physical problems in devices caused by impacts or not being properly oriented.

In order to identify failures and errors in ad-hoc mobile networks and wireless mesh networks, in 2007 the authors of [SELF-38] proposed a novel self-diagnosis model called "Adaptive-DSDP". This protocol is based on comparison where tasks are assigned to pairs of nodes and the results obtained are analysed and compared.

Cheng and Tsai [SELF-39] developed a day and night traffic surveillance system with the ability to self-diagnose whether or not vehicle tracking should be performed based on road lighting and weather conditions. If so, the system tracks the vehicles one by one and generates current traffic parameters. In the negative case, an estimation of parameters is carried out by means of regression to estimate the number of vehicles that circulate through a specific sector during a defined time. This self-diagnosis increases the reliability of the system.

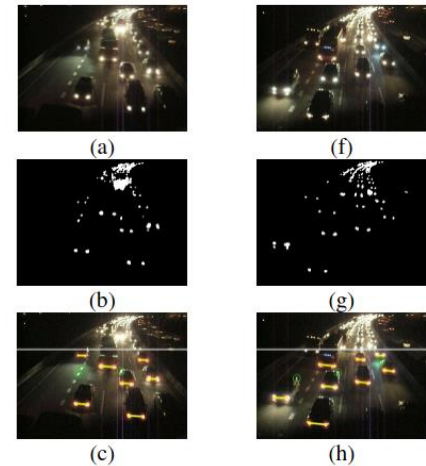


Figure 32. Examples of vehicle detection in the traffic surveillance system [SELF-39].

3.2.2.3.4. Self-healing

Self-healing is a part of autonomous systems that is responsible for independently managing the recovery of the parties affected by a failure or attack without human intervention. This mechanism provides the ability to maintain and resume the system in an automatically set condition [SELF-40]. Khalil et al. in [SELF-41] also include failure detection as part of self-healing. In [SELF-10], IBM explains that self-healing is the self-capability to automatically diagnose and resolve, both, hardware and software failures.

In China, Yang et al. [SELF-42] developed and implemented a self-healing system for the electrical network made up of Easergy T300 controllers installed in medium voltage feeders (20.000 V) that monitor the state of the electrical network through an analysis and self-healing algorithm in real time to detect failures and avoid prolonged power outages. The controllers analyse the load of the feeders, obtaining data on the temperature of the devices, energy, etc. in order to manage the network. Thanks to the self-healing algorithm, the system is capable of identifying the type of fault and its location, isolating the sector of the network with problems and reconfiguring the network to re-energize the areas affected by the fault. In this way, the duration of power outages can be reduced from hours to just seconds autonomously.

In [SELF-43], the authors also develop an autonomous control system for the monitoring and self-healing of the smart distribution network based on distribution automation and advanced distribution automation. The self-healing of the system includes the preventive self-healing, the fault self-healing and the economical self-healing. This intelligent system is capable of adapting to the complex environment formed by these networks, continuously monitoring and managing resources. Thanks to this, the system is able to ensure and improve the electrical supply of the network in the event of a problem thanks to the use of resources such as power generators widely distributed throughout the network, energy storage devices and even electric vehicles connected to the network (V2G).

Control of autonomic systems through monitoring their health status is one of the essential parts of self-healing algorithms. [SELF-40] proposes a monitor model that can improve self-healing performance by decreasing the amount of resources spent on self-healing affected parts of the system.

Neural networks are complex algorithms used in a wide variety of applications [SELF-41], especially in the field of artificial intelligence. To avoid failures in these systems, there are self-healing algorithms that are based on replacing defective hardware nodes with new ones, which causes system overloads [SELF-41]. Khalil et al. [SELF-41] propose a novel method that using a single node per layer it is possible to replace any defective node.

If a node fails, its neighbour will also perform its tasks (apart from those already assigned to it) sequentially. If the neighbouring node fails, the only spare node will take over, reducing the load on the system.

Liu et al. [SELF-44] show the design and implementation of a zero-time self-healing communication network for real-time ship monitoring. This network is capable of connecting sensors, control devices and computers to interact with the ship's maintenance team. Through various control and surveillance mechanisms, it is capable of automating many of the tasks carried out on ships. The objective of this novel design is to solve the transmission, reliability and real-time problems of network communications. To do this, it transmits the information through several routes to have a seamless and instantaneous self-healing network. Thanks to this network, the maintenance of the ship becomes easier and faster.

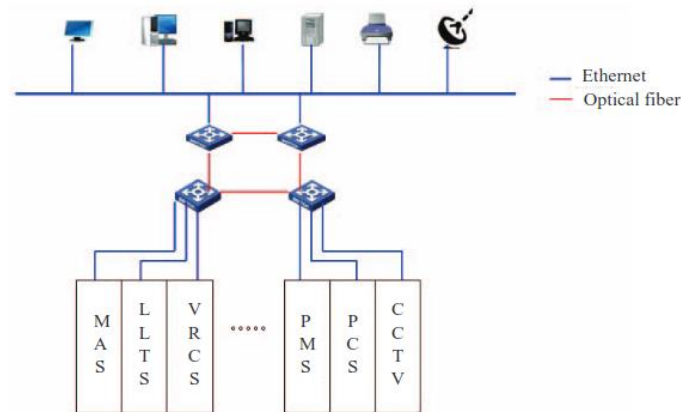


Figure 33. Communication network for real-time ship monitoring [SELF-44].

In [SELF-45], as in [SELF-42] and [SELF-43], the author exposes a model for the automatic reconstruction of the electrical network with self-healing capacity to avoid power outages to users and reduce the cost of repairing the electricity network.

3.2.2.3.5. Self-scaling

Based on the definition offered by Herbst et al. [SELF-46] on scalability, we can define self-scaling as the self-capability of an intelligent node to increase or decrease the use of its resources depending on the volume of work to be done. If the workload increases, the node is able to increase its resource usage automatically. Otherwise, it will remove part of its resources to accommodate the volume of incoming work.

Herrera and Moltó [SELF-47] introduce two novel biology-inspired algorithms that enable auto-scaling in architectures based on the execution of self-managed containers. The algorithms described are:

- Self-scaling self-sufficient cell model (SCM): this model is characterized by the lack of direct interactions between containers. This design, in turn, is subdivided into 3 variants (SCM-A, SCM-B and SCM-C).
- Self-scaling interactive cell model (ICM): this model is characterized by containers that have information about the containers that are in their environment. The exchange of information can be done directly (between containers) or through intermediate services.

In [SELF-48] the authors describe a model for self-scaling the resources of a network based on the task execution times of each instance of virtual network functions (VNF). The resources used by each instance (both physical and virtual) are assigned per cycle unit using a weighting factor. The system is made up of two components: a self-scaling application (which includes several control and management modules) and a monitoring module based on micro-services. Nikraves et al. [SELF-49] propose an architecture for a self-scaling prediction ensemble based on empirical studies, which is capable of selecting the best prediction algorithm based on the amount of real-time workload.

Casalicchio and Perciballi [SELF-50] present a self-scaling model called "KHPA-A" that connects to the Kubernetes controller and is based on a type of metric called absolute. This algorithm can make use of the input parameters used by the original "KHPA" algorithms to obtain the number of containers to be instantiated. The

use of this type of metric allows the system to reduce the response time of the applications compared to the current Kubernetes self-scaling algorithm.

Chattopadhyay et al. [SELF-51] propose a self-scaling orchestration model for IoT applications called "Aloe". This framework dynamically deploys lightweight controller instances close to IoT devices (which are resource constrained) to ensure high availability and low setup time. It is fault tolerant, can migrate instances from one site to another in case of problems with part of the network, and uses Docker as a base to support migration.

3.2.2.3.6. Self-configuration

According to [SELF-52], the self-configuration of an application or autonomous system is the self-capacity of being able to configure and reconfigure itself automatically and independently in any type of possible condition. In [10], IBM explains that self-configuration is the self-capability of autonomous systems to configure themselves and their components following high-level policies.

Yang et al. [SELF-53], in 2010 developed a model to self-configure connected terminals in 4G networks and heterogeneous communication and service environments. When a terminal connects to the network, the framework puts it in pre-operational mode until the node self-configures, at which point the node becomes operational within the network. When a terminal leaves the network, the TMS (Terminal Management System) notifies the rest of the nodes so that they are aware of the new state and reconfigure themselves appropriately. Wang and Vanninen [SELF-54] describe and compare different protocols for individual peers to self-configure the P2P network. To determine which is the best protocol, they simulate small-scale P2P networks and compare the quality of self-configured networks.

In 2020 Mombello et al. [SELF-55] presented a self-configuring system for a photodetector sensor. Its goal is to use a control unit that can be programmed to find the centre of the light beam hitting the sensor, and then set the detection pattern. This model allows you to automate the alignment of the light beam with the detection pattern. For this, the model is capable of obtaining data from the light sensor to reprogram the behaviour of the photodetector sensor in real time. In [SELF-56], the authors describe a self-configuration algorithm for a modular robotic system (MRS). This system is made up of robots which move through a virtual grid until they reach their optimal position in the configuration space. Through local communications the robots can analyse and plan routes within the grid to change position.

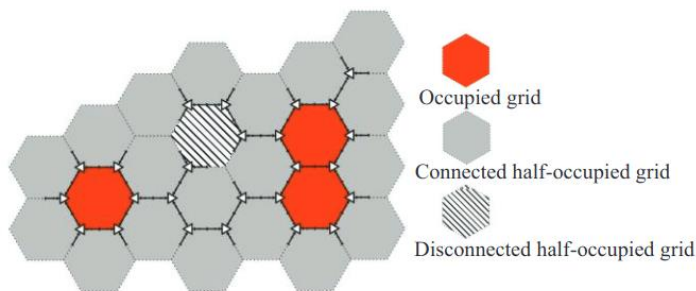


Figure 34. Example of system grid partition [SELF-56].

Currently, there are millions of applications running that offer services to users. In order to be updated, many of them must be taken off-line, their components updated, rebuilt and reconnected. This method leaves users temporarily without service, and there are systems that cannot afford these complete interruptions, only partial ones. Abdellaoui et al. [SELF-57], propose a real-time self-configuration system that is capable of automatically connecting and disconnecting the modules (components) that make up the

applications to reduce service outages and cause the least possible interruptions. Each connected object in the application is considered as a software module that is added or removed to be updated separately.

Yao et al. [SELF-58] have designed a system that automates the self-configuration of the use of virtualised shared resources in graphics cards of cloud servers intended for cloud-gaming. This framework is made up of four modules:

- Sensor module: gathers preliminary system and application data.
- Modelling module: automatically analyse raw data from the sensor module.
- Controller module: for each virtual machine running on the graphics card, an agent monitors its performance and sends the information to a scheduler. This analyses the information of all the virtual machines and sends an instruction to activate the control system.
- Self-control-configuration module: manages the self-configuration of the controller parameters.

In [SELF-59] the authors present a novel self-configuration model, based on software-defined networks (SDN) for time-sensitive networks. In existing configuration methods, the end nodes have to send their data to a central management node. These methods require manual configuration of the hosts. The new algorithm allows resources to be obtained in a transparent and automated manner, facilitating self-configuration in heterogeneous environments.

3.2.2.3.7. Self-optimisation

In 2003, IBM listed self-optimisation as one of the four basic pillars of an autonomous system. IBM defined the concept of self-optimisation in autonomous computing as the continuous improvement of the performance and efficiency of an autonomous system [SELF-10]. For Nami and Bertels [SELF-60], self-optimisation is the ability of an autonomous system to allocate resources and use them in the most efficient way possible, meeting user requirements. In addition, they also state that autonomous system workload management and resource usage are two important points in self-optimisation.

Zheng et al. [SELF-61], defined a model based on autonomous computing to automatically optimise services offered to users. When the system changes internally, that is, the parameters that influence the performance of the services provided to users change during its execution, dynamic self-optimisation is executed. This improves the performance of the service to make it more efficient. When there are no big changes internally in the system, the static self-optimisation prediction is executed. Both methods are combined to automatically improve the performance and efficiency of the services that the system provides to users.

The authors of [SELF-62], propose a method to automatically optimise handover parameters for 5G networks. In these networks, configuration of handover control parameter (HCP) settings is done manually or through self-optimisation functions. Due to the large number of devices connected to the network, offering a stable connection for all over time has become one of the priorities in this type of network. Device handover occurs when a node moves between two cells of a network. The authors also classify the current algorithms as central optimization models, that is, the optimization is performed based on the performance of the network as a whole and not individually for each connected device. To change these behaviours of the network, in the paper [SELF-62] the authors describe a handover self-optimisation technique for each user independently. To do this, the algorithm predicts the HCP configuration for each user based on a weight function.

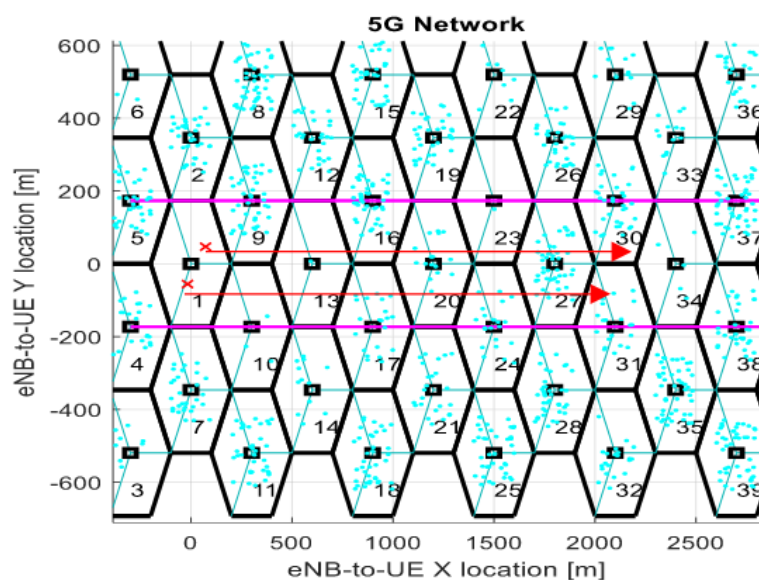


Figure 35. 5G network model with hexagonal cells, divided into three sectors [SELF-62]

In [SELF-63], Sánchez-González et al. propose a rule-based self-optimisation model for mobile networks that improves and speeds up convergence in the search for solutions. These rules are really information on how to solve specific problems. In addition, the authors state that this system has been fine-tuned to improve coverage and cell overlap within the same network. Trumler et al. [SELF-64], presented a model for creating self-organising autonomous systems that are based on nodes located in the network. This system employs a mode of operation based on the hormonal system of humans. Each node sends information for self-organisation

through messages without using any extra communication system to avoid overloading the network. The objective of these messages is to know the consumption of the resources of the nodes to be able to optimise them in the most efficient way. The algorithm works in conjunction with a middleware also developed by the authors of the paper.

In [SELF-65], the authors implement a self-optimisation model for the nodes of cognitive wireless home networks, called “Home Cognitive Resource Manager” (HCRM). The system uses several self-optimisation algorithms and information captured from the execution environment in order to perform efficient radio resource management. To achieve its goal, the framework uses utility-based reasoning and compliance with policy regulations.

Wang et al., describe in the paper [SELF-66] an autonomous system for self-optimisation of the course of a ship. To do this, the objective to be achieved by the system is established and, through various algorithms, it determines the most optimal and efficient control parameters of the ship's course.

3.2.2.3.8. Self-adaptation

Self-adaptation is the self-capability of the autonomous systems to adjust their behaviour during execution in real-time. This adaptation is made to respond to changes in the perception of its environment and of the system itself [SELF-67][SELF-68].

Amiri et al. [SELF-69], propose an autonomous system that uses a dynamic router architecture capable of adapting at runtime. Several studies by the authors of the paper indicate that centralised routings offer greater reliability, while decentralised ones offer more performance. This system performs multi-criteria analysis to optimise and self-adapt the architecture between more centralised or more distributed routing to deliver the highest reliability and maximise performance.

The work described in [SELF-70], deals with the variation of the Particle Swarm Optimization (PSO) algorithm with dual self-adaptation and dual variation to improve the premature convergence problems of the standard version. The goal is to widen the search range for the optimal solution and improve the search accuracy, the algorithm's rate of convergence, and its response speed. The authors affirm that applied to the optimization of objective functions, their version of the PSO improves performance and results compared to the standard version.

On the one hand, in [SELF-71] the authors describe a multi-tier self-adaptation model for microservice systems that aims to improve the self-adaptation capabilities of microservice frameworks. In addition, they also present a self-adaptive description language with which to determine the adaptation logic at the different levels of microservice systems and a platform called “AdaptiveK8s” to provide support as a Kubernetes extension. The goal of all these efforts is to specify self-adaptation requirements at the different levels and to provide the necessary components to improve self-adaptation in microservice systems. On the other hand, Nallur and Bahsoon [SELF-72] propose a decentralised model in the cloud that uses heuristics so that service-based applications can self-adapt at runtime to the quality of service requirements they offer to users.

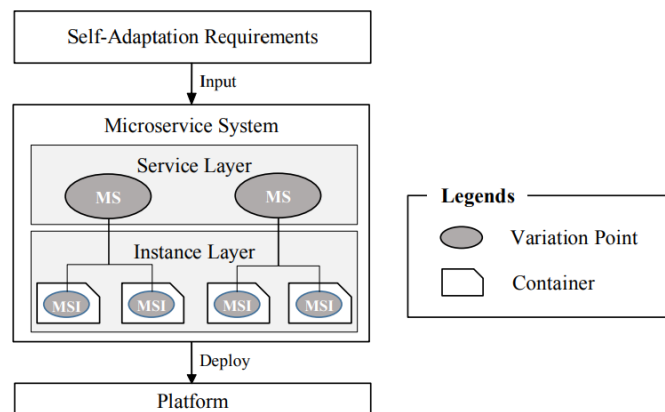


Figure 36. Self-adaptation model for microservice systems [SELF-71]

Ardito [SELF-73], developed a system to self-adapt the operation of smartphone applications in real-time depending on the current battery consumption of the device. The goal is to reduce the energy consumption of smartphones and extend the life of their batteries. The method has several phases of operation. First, the power management module of the operating system obtains the consumption values through the hardware. Second, the module analyses and divides the energy expenditure between each running application based on the current use of each one. Finally, it sends the information with a maximum threshold that must not be exceeded. If the

application exceeds the threshold, the operating system sends it a warning to modify its operation, adapting itself according to its energy consumption.

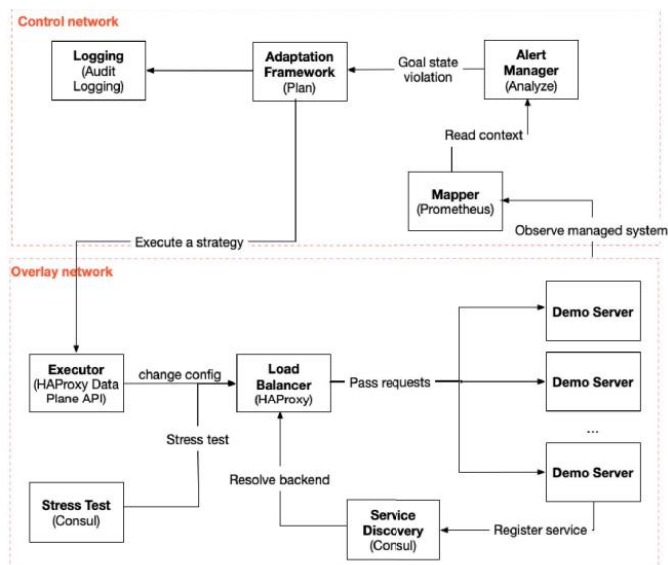


Figure 37. Networks of the self-adaptive system [75].

In [SELF-74], Yuan et al. present a self-adaptive model called "CASC", based on MAPE [SELF-24], to adapt the composition of services in real-time. Self-adaptive composite services can automatically adjust in real-time to changes in their surrounding environments. This system is capable of self-adapting by selecting new services or generating new schemes for the composition of the service.

Boyapati and Szabo [SELF-75], developed a self-adaptive system for large-scale microservice architectures, based on MAPE-K [SELF-24]. The system is made up of two independent networks. In a network, the MAPE-K loop monitors the environment, analyses the information received, and schedules tasks. On the other network, the scheduled tasks are executed on the managed system. All components are deployed on Docker and are related to each other through REST API.

The authors emphasise the use of open source

tools for the development and implementation of the proposed system.

3.2.2.3.9. Self-learning

Based on [SELF-76], we can define self-learning as the self-capability of an autonomous system to improve its performance using unsupervised artificial intelligence and machine learning over time.

Dongzhi et al. [SELF-77], developed a self-learning system for fault diagnosis based on an ontology knowledge data store. To achieve a correct diagnosis of failures, the model is capable of drawing conclusions from the knowledge stored in the data warehouse. This new information is stored in the fault diagnosis ontology to adjust the knowledge of the database and automate and improve the process of diagnosing system faults. In [SELF-78], Zhang et al. created a controller based on the self-learning of parameters for the propulsion system of an electric vehicle in order to improve acceleration from a standstill and speed recovery while driving.

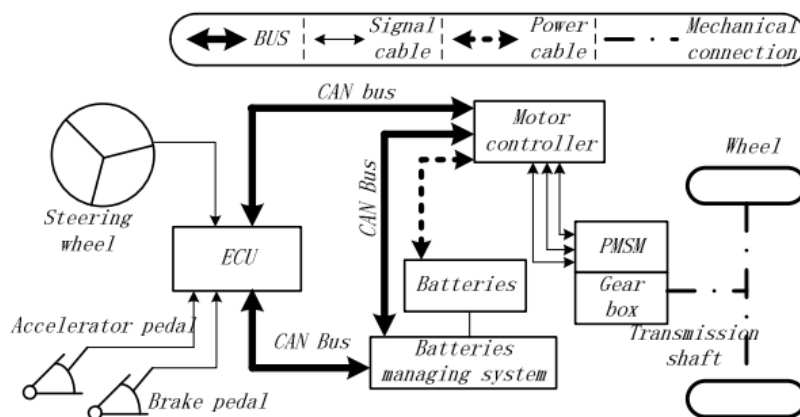
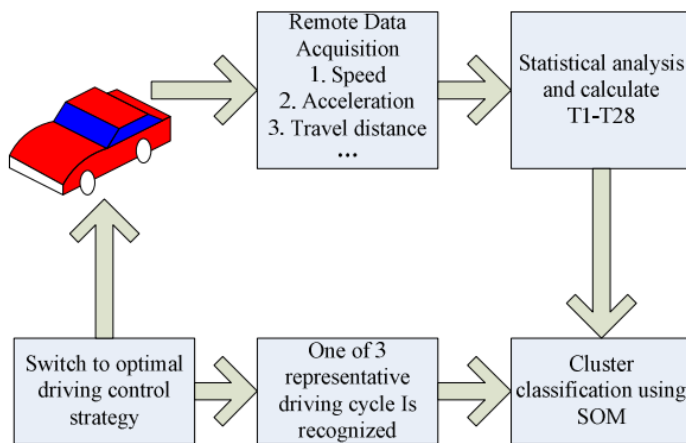


Figure 38. Propulsion system of an electric vehicle [SELF-78]

Wen-Bin [SELF-79], in 2012 proposed a model with self-learning of parameters to control the temperature inside a spacecraft constantly. The objective was to guarantee that both the components of the ship and the control and work systems, as well as the living beings that lived inside, all had the correct temperature. To achieve this goal, the temperature control parameter is capable of self-learning from the ambient temperature. Unlike direct on/off temperature control systems, this algorithm is capable of continuously modifying the amount of time the heating system has to be on and its intensity (power) to maintain the temperature of the object controlled in the optimal range.



Jamshidpour et al. [SELF-80], implemented a system based on self-learning for the classification of high-resolution hyperspectral images. The system works with two learning algorithms, active learning (AL) and semi-supervised learning (SSL). The SSL algorithm semi-tags those untagged samples. The best labels are added to the training set for the classifier. The AL algorithm also selects unlabelled samples and adds them to the training set but classifies them by human experience. The advantage of this system is the reduction of human interaction and the ability to vary the degree of involvement of each algorithm.

To increase the efficiency and decrease the energy consumption of electric vehicles, Ji-Hui et al. [SELF-81]

Figure 39. Remote self-learning driving cycle [SELF-81]

developed a self-learning algorithm of the driving cycle for these vehicles. This system, by means of a device connected to the car via the CAN bus, obtains up to 28 different parameters on the driving of the vehicle and its status and sends them wirelessly to a data server via the Internet. Based on the result obtained after analysing the data, the driving cycle is modified to optimise the vehicle's efficiency. With each cycle of analysis of the information received, the self-learning system is able to more efficiently optimise the vehicle's control parameters.

Chen and He [SELF-82], applied in 2016 for a patent for an intelligent system that saves electricity in drinking water dispensers in offices and homes. This ingenious system is made up of hardware components (composite control switch, time switch, control unit, etc.) managed by a self-learning fuzzy control algorithm to reduce unnecessary heating of water. The model has four parameters: initial heating time (T1), reheating time (T2), reheating time if the water is not removed (T3) and time between two reheatings if the water is not removed (T4). Depending on the season of the year, the outside temperature, the use, etc. these parameters vary constantly. However, the system, through its self-learning module, is able to determine the best initial combination to efficiently heat the water and save energy.

In [SELF-83], Abeysinghe and Bandara present a new self-learning algorithm to detect the incompatibility of opinions in the TripAdvisor travel social network. The model, applied to hotel reviews on the social network, is capable of finding inconsistencies between the opinions written by users and the bubble ratings extracted from the reviews. These bubbles define the overall experience of the place. The system is capable of reliably determining correct reviews and ratings, to train the self-learning method to detect reviews that do not match the ratings using the matching reviews. One of the novelties of the algorithm is the use of three types of ratings (positive, neutral and negative) instead of the common two (positive and negative).

3.2.4. Data syntactic and semantic interoperability in the continuum

We are currently witnessing an exponential growth in the number of solutions that offer/process ever-increasing amounts of diverse data. To some extent, this is related to the proliferation of “sources” that are various types of IoT devices, but also to a significant increase in the number of solutions and systems whose applications are already entering virtually every area of our lives. A significant number of “data producers” still offer data either in an unstructured form (e.g., most IoT devices) or using dedicated, often proprietary, data formats. Since aerOS, through its modular architecture, will provide the basis for a scalable, decentralized and adaptable computational continuum, one of its core components needs to offer extensive support for data-level interoperability, considering many aspects of data sharing/processing. For simplicity, let us treat all the devices, services, systems, etc. that aerOS can manage uniformly, and name them *artifacts*.

3.2.4.1. Interoperability levels

Although interoperability is a relatively complex concept, with many aspects to consider, existing “generic” definitions, boil down to the observation that interoperability is the ability of two or more artifacts to work together despite differences in language, interface, or execution platform [DIC-1]. Literature offers several classifications, known as “levels of interoperability,” taking different aspects of the notion into account. One of the most popular ones is the LCIM (Levels of Conceptual Interoperability Model) [DIC-2] classification. It consists of seven *levels*, named L0 to L6, ranging from no interoperability at all, to conceptual interoperability. The classification, originally created in the context of simulation theory, provides/recognizes the following levels:

- Level 0 – *No interoperability*.
- Level 1 – *Technical interoperability*. Artifacts have technical connection(s) and can exchange data between themselves. The premise are common communication protocols (such as HTTP, TCP/IP, UDP/IP etc.) and the network connectivity.
- Level 2 – *Syntactic interoperability*. Artifacts have an agreed protocol to exchange the right forms of data in the right order, but the meaning of data elements is not established. The contents clearly defined are the format of the information exchanged (XML, SOAP, JSON, etc.).
- Level 3 – *Semantic interoperability*. Interoperating artifacts are exchanging a set of terms that they can semantically recognize. The information defined are the meaning of the data and the content of information exchanged.
- Level 4 – *Pragmatic interoperability*. Artifacts are aware of the context (their states, processes, etc.) and meaning of information being exchanged. The information defined are the use of the data and the context of information to be exchanged.
- Level 5 – *Dynamic interoperability*. Interoperating artifacts can re-orient information production and consumption based on understood changes to meaning, due to context changes over time.
- Level 6 – *Conceptual interoperability*. The interoperating artifacts are completely aware of each other’s information, processes, contexts, and modeling assumptions. The level is focused on the composability and the abstract modelling of the domain.

Another, more compact, classification has been proposed as a part of the European interoperability framework for Pan-European e-government services [DIC-3]. It recognizes three levels of interoperability: *technical*, *semantic*, and *organizational*. Yet another classification proposal, provided by ETSI and AIOTI [DIC-4], defines four levels: *technical*, *syntactic*, *semantic*, and *organizational*, where some technical-level aspects have been moved/separated into a new category (syntactic interoperability).

- *Technical interoperability*. It is usually associated with artifacts, that enable machine-to-machine communication to take place, mostly communication protocols and the infrastructure needed for those protocols to operate. Some protocols in common use include: CoAP, HTTP, WebSocket, MQTT and AMQP.
- *Syntactic interoperability*. It is usually associated with data formats. The messages transferred by communication protocols need to have a well-defined “syntactic” representation.
- *Semantic interoperability*. Refers to the meaning of data and concerns the human rather than machine-level interpretation of the data. Thus, interoperability on this level means that there is a common understanding of the meaning of data being exchanged between artifacts.
- *Organizational interoperability*. Refers to an organization’s ability to effectively communicate and transfer (meaningful) information (data) despite the fact that they may use many different information systems, as well as operate under different geographic or cultural conditions that can have a significant impact on their operations.

From the technical, “data layer” point of view, taking the ETSI classification as the reference, we shall concentrate on two aspects – the *syntactic* and *semantic* interoperability.

3.2.4.2. Data-centric interoperability

Syntactic interoperability involves the use of common data formats and common data structure protocols. It is a necessary prerequisite for the existence of semantic interoperability, enabling and facilitating data sharing and processing. Semantic interoperability, on the other hand, refers to the ability of artifacts to exchange and process data based on a uniquely defined common meaning/interpretation.

Since aerOS shall offer support for computations ranging over the entire *edge-cloud continuum*, in particular, it will need to utilize/provide interoperability mechanisms starting from the “low level” data producers, i.e., IoT devices. Here, due to the rapid technological development, there is still some lack of standardization. In particular, there are no “official standards” for the syntactic representation of data produced/utilized by various types of IoT devices. Fortunately, in most cases, data in “proprietary” formats can be converted into one of the commonly used representations at a relatively low (computational) cost. Therefore, the task of achieving the syntactic interoperability is generally well understood and, thanks to the existence of popular, well documented and widely supported data formats, such as XML [DIC-5] or JSON [DIC-6] relatively straightforward to accomplish.

In contrast, the task of achieving semantic interoperability is much harder, since it requires machine interpretable (and “understandable”) semantic descriptions. Such descriptions, for example, in the form of:

- data models and data types,
- models describing the interaction with artifacts,
- frameworks for describing different versions of artifacts,
- semantic descriptions of artifacts and the context,
- privacy and security policies covering use of data,
- smart contracts and terms & conditions

can be further utilized to establish semantic interoperability solutions.

To minimize barriers for digital services that span different platforms, there is a strong need to encourage convergence on modelling frameworks and languages. Some relevant work that can be considered includes:

- Resource Description Framework [DIC-7] (RDF) – using graphs with directed labelled arcs to represent information.
- JSON-LD [DIC-8] – JSON-based serialization of RDF, using JSON Schema [DIC-9] for describing the data types.
- Web Ontology Language [DIC-10] (OWL) and RDF Schema [DIC-11].
- Entity Relationship Diagrams [DIC-12] (ERD) and Unified Modelling Language [DIC-13] (UML).

The most general yet promising approach to the problem of semantic description of artifacts and the data exchanged between them seems to be the use of technologies developed so far for the Semantic Web [DIC-14]. This includes application of languages such as, mentioned above, RDF and JSON-LD for representing “semantically annotated” data. These, in turn, require existence of appropriate semantic model descriptions in the form of *ontologies*, which are sets of objects and relationships used to define and represent given area of concern. Ontologies provide an abstraction which aims to hide heterogeneity of artifacts and enables them to exchange and process data with meaningful content, thanks to “semantic annotations” based on ontologies. Ontologies, which are the necessary component of the solution, can be formally defined using languages such as RDF Schema (RDFS) and (restricted “variants” of) OWL.

Defining ontologies requires a certain amount of expertise, or at least knowledge of the modeling language one wants to use. In some cases, however, you can use tools such as Ontomizer [DIC-15] or ReDeFer [DIC-16] to automatically generate an ontological model. In the case of XML, such a model can be obtained from a data structure definition expressed, for example, in the XSD language, or even “raw” XML data. Of course, the quality of the resulting ontology may leave a bit to be desired.

The choice of a particular set of ontologies will, of course, depend on the application area. However, given the nature of the edge-cloud continuum environment in which aerOS instances will operate, IoT-related solutions will play an important role. Therefore, proper semantic treatment of sensors, sensor networks, actuators, and their operations, i.e. observations and actuations, will be of fundamental importance. Many ontologies for describing these concepts/entities have been proposed. Their overview can be found in [DIC-17]. More recently, as a result of the INTER-IoT project, two modular ontologies, targeted at IoT platforms deployments have been proposed [DIC-18].

Usually, for obvious reasons, the IoT-dedicated ontologies would also be combined with or utilize other, high-level or domain specific ontologies. Among high-level ontologies, there are “top-level” ontologies, modelling very general concepts, that are common across all domains [DIC-19]. Other types of “general ontologies” would typically be ontologies representing widely used concepts/domains, such as geolocation (e.g., LinkedGeoData [DIC-20], GeoSPARQL [DIC-21], or WGS84 [DIC-22]), units of measure (e.g., QU [DIC-23], OM [DIC-24], or SWEET units [DIC-25], time (e.g., Time OWL [DIC-26]), or provenance (e.g. PROV-O [DIC-27]). A useful survey of data management and integration related ontologies can be found in [DIC-28].

In most realistic applications, to achieve interoperability of artifacts at the data level, it is not enough for the data to be semantically annotated. In general, artifacts working together within an ecosystem will not use a unified semantic model. Therefore, it becomes extremely important to be able to “translate” data while preserving its meaning, expressed in the semantics of the sender and receiver, respectively. Most of the data produced and processed in the environments that aerOS will support is streaming in nature. General stream processing solutions, such as offered by Apache projects – Kafka [DIC-29], Storm [DIC-30], and Flink [DIC-31] offer modern tools for stream data processing. However, they do not provide direct semantics handling capabilities. An interesting, scalable, and highly efficient solution to this problem was provided by the INTER-IoT project, through the Inter-Platform Semantic Mediator tool (IPSM) [DIC-32]. In its implementation, IPSM utilizes Apache Kafka, and offers high-performance, scalable and highly configurable semantic translation mechanisms. Recently, a general RDF end-to-end streaming solution, named Jelly has been proposed [DIC-33]. It is simple to implement/utilise, flexible and applicable to wide variety of use cases. Jelly can be considered as an efficient complement to the semantic translation mechanisms offered by the IPSM platform.

3.2.5. Data sovereignty, governance and lineage policies

3.2.5.1. Overview

Data governance is defined as “a data management function to ensure the quality, integrity, security, and usability of the data collected by an organization” [DSGP-1]. Expanding the definition, data governance seeks to engage people, processes, and technologies to maximize the value of data while preserving privacy and protecting access to the data [DSGP-2].

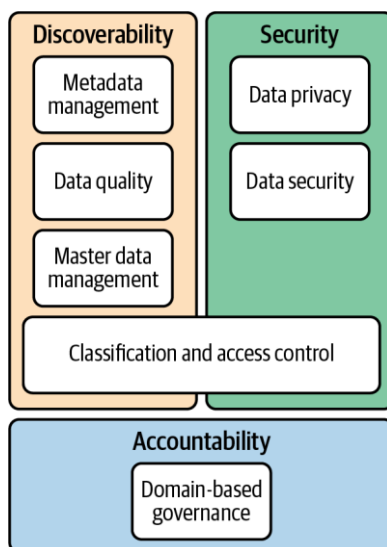


Figure 40. Main aspect of data governance [DSGP-1]

The purpose of data governance is to build trust in data. To this end, the data governance strategy addresses three key aspects, as captured in Figure 40:

- **Discoverability**

Data consumers should have easy and reliable mechanisms for finding the right data within the organization. It is essential for them to know the location of the data, its meaning, and potentially, its relationship with other data assets. In this sense, metadata enables the discovery and governance of data across the organization. Metadata can be automatically generated by crawling data sources and monitoring data pipelines, but also it can be manually curated by humans. In this sense, several technologies like metadata management tools or data catalogues have emerged to facilitate the collection and exposing of metadata [DSGP-3] – [DSGP-6].

Metadata can be divided into four main categories: business, technical, operational, and reference. Business metadata relates to the meaning of data and how it is used by business applications, i.e., the definitions to properly use data. Technical metadata describes technical details on data during its lifecycle such as data models, field mappings, data lineage, or workflow orchestration. Operational metadata helps at monitoring the processes involved in the lifecycle of data by providing information such as runtime logs, statistics, or job IDs. Reference metadata helps classifying data based on standard references that can either internal or external.

Other key dimension of discoverability is data quality. Data quality is the process that optimizes data by making sure it meets a set of requirements such as accuracy, completeness, and timeliness.

Lastly, master data management focuses on the definition of consistent entities across the organizations. This harmonization guarantees the proper classification of data, which then facilitates the application of data protection and data access policies.

- **Security**

Organizations must handle data in conformance with regulations (e.g., GDPR) along with a careful management of sensitive data (e.g., PII). On the other hand, data access policies and protection systems must be prepared to protect against threats like data leakage or unauthorized data access.

- **Accountability**

Assigns people responsible to govern a fragment of data. These people are also known as owners or governors. Data owners can be responsible of data at different levels of granularity, being the domain-level the current trend with new approaches like data mesh.

3.2.5.2. European Data Governance Act

Data keeps growing at an unprecedented pace, and unlocking its full potential is essential for driving innovation in industries and SMEs. A correct and efficient processing of data can help in developing new services and products which citizens can benefit from. But to correctly exploit data rules and measures must be applied to guarantee trust in the data.

As part of the European Commission's Strategy for data the Data Governance Act (DGA) provides guidelines for increasing trust in data, improving the mechanisms for accessing data, and promoting data reusability. This initiative seeks to involve parties both from the public and private sectors with the goal of creating and developing ecosystems for data sharing across the EU in strategic domains such as energy, health, or mobility.

3.2.6. Advanced AI management approaches

AI activities require a lot of computing and are typically trained, developed, and used in data centers with specialized servers. However, with growing power of mobile devices, significant number of intelligent applications are anticipated to be implemented at the edges of cellular connections [AI-1] benefiting from the concepts of Internet of Things, edge-cloud continuum [AI-2]. AI at the edge of the network promises to be beneficial not just at functional but also at business level, allowing the realisation of federated/distributed AI scenarios and adjusting to the capabilities of the continuum applying techniques such as frugal AI. Moreover, processing data close to the edge of the network can reduce latency and improve privacy by eliminating data sharing.

3.2.6.1. Federated Learning

Federated Learning (FL) is an approach to machine learning, where training of a model involves multiple datasets stored in “local nodes” (clients), while the training itself proceeds without exchanging any data. In other words, there exists a central (shared) model whose sub-versions are trained in each participating node using only its local data. Next, model parameters are “combined into the central model”. After the update is completed, the updated central model is redistributed (back) to the nodes that participated in the training. Here, the loop closes, and the process repeats, until the common model is considered to be of “good enough quality” (using process specific criteria).

The rationale behind FL is: rather than splitting data of a single stakeholder, and training the model in parallel, the focus is on the use of local data that may belong to different stakeholders (without sharing it). It is easy to see that only local nodes have actual access to their own data (which is not shared), while the central (shared) model is updated on the basis of results delivered by individual nodes.

In the articles [AI-3, AI-4], the foundations of FL and an architecture of the FL-based system, have been proposed. The used machine learning software was TensorFlow, while the application area was related to word prediction and suggestion to be provided as a service when smartphone users are typing “messages”.

Designing an FL-based system requires addressing problems that are (typically) not present in other popular ML approaches [AI-5]. These problems are mostly related to the fact that data is in multiple locations which has consequences: (1) cost of communication needs to be considered, and (2) data can be unbalanced in a serious way, different from the situation when data is not distributed.

The article [AI-6] focuses on the problem related to the communication between the FL participating nodes and the global server. Specifically, authors propose an encoding, which allows to reduce the size of transferred data by up to 32 times. This is especially applicable in a situation where local nodes finish the work in a similar time and send updates to the server causing congestion.

One example of an FL application is described in an article [AI-7], i.e. a solution for classifying signals from the electroencephalogram (EEG). Here, due to the need for personal data protection, multiple small data sets exist that, due to privacy policy, cannot be combined into one large (training) data set. An algorithm using the method of covariance, based on neural networks, has been proposed. The article describes how signals are processed to constitute an input to the neural network. Next, an averaging method is applied, and the model is updated on its basis. The achieved results are satisfactory, compared to other algorithms.

The authors of [AI-8] describe their FL solution for processing medical data. The proposed approach is tested using the MIMIC-III database. The authors do not describe, in detail, the used algorithms, but only the components that they consist of. The client consists of three parts: the first for training, the second for communicating with the servers, and the third for performance testing.

Another healthcare application example is described in the NVIDIA blog article. The Nvidia Clara platform is used to implement the proposed approach, i.e. an FL capable platform designed to process medical images and genomes. The main motivation behind the described approach is, again, protection of patient privacy. Servers in hospitals train the global model on local data. Local results are sent securely to the global server.

In [AI-9] an FL-based solution for keyword detection is described. The model uses the encoder-decoder architecture. A modified FedAVG algorithm was used, in which the Nesterov accelerated gradients were used for the server-side updates. Various methods of server-side optimization were also compared, inducing Adam, Yogi and LAMB.

In [AI-10] a system supporting maintenance of industrial machines is described. Normally, machine learning is based on local data available within each machine. Use of FL, allows one to benefit from the data of business partners, without the need to share the actual data. Additionally, FL applied in this case requires appropriate data preparation e.g. handling interoperability.

The publication [AI-11] discusses the possibility of combining FL with Generative Adversarial Networks (GANs) that are used to generate elements in various categories. They consist of two components: a generator and a discriminator. The generator learns how to create elements like a given category of “objects”. The discriminator, on the other hand, learns to distinguish between true (correct) and false (incorrect) “objects”. In

the described system each client has a generator and a discriminator module. Clients also update the global (shared) generator and discriminator, located on the server. It is claimed that, due to the system consisting of two modules, the problem of model synchronization is more complex. Four methods of synchronization are discussed: synchronization of the generator, the discriminator, both elements, and lack of synchronization are considered and compared. For the real-world cases, where communication costs are very high, it is suggested that generator-only synchronization should be used. In other cases, use of synchronization of both generator and discriminator is proposed.

Table 1. AI tools for Federated Learning

Tools for Distributed AI	
Caffe2	Deep learning framework with multiple algorithms merged into PyTorch API
CNTK	The Microsoft Cognitive Toolkit - cloud-based deep learning framework (not longer actively developed)
DIANNE	Modular ML framework for designing, training and evaluating artificial neural networks
TensorFlow	Google ML framework
Single-Machine ML Systems and Libraries	
Theano	Python library and compiler to optimize math calculations
Caffe	(Convolutional Architecture for Fast Feature Embedding) deep learning framework
SciKit-learn	Python ML library featuring various classification, regression and clustering algorithms
PyTorch	ML framework based on Torch library under the Linux Foundation umbrella
MLPack	Machine learning software library for C++
NVIDIA libraries	Rapids, cuBLAS, faster ML training
Tools for Federated Learning	
PaddleFL from Baidu	Open-source federated learning framework based on PaddlePaddle: Python, C++, GPU support library
Flower	Federated learning framework: customizable, extendable, framework-agnostic (can be used with e.g. PyTorch, TensorFlow, MXNet, scikit-learn)
Google TensorFlow Federated	Open-source Python federated learning, TensorFlow based framework created by Google
Threepio - PyTorch, Tensorflow.js, and TensorFlow	Javascript library enabling to run visual training with TensorFlow
IBM Federated Learning	Proprietary Python framework with large number of implemented ML algorithms to build FL systems supporting Keras, PyTorch, SkiKit-learn and TensorFlow
Federated Core	Programming environment for implementing distributed computations, tensorflow federated
Federated AI Technology Enabler (FATE)	Open-source project initiated by Webank's AI Department - distributed Python framework with Docker, k8s aligned to big data
KubeFate	Environment for distributed and federated learning using docker and k8s with Python Spark
Fate Cloud	Cloud infrastructure working with KubeFate
OpenMined PySyft	Python federated learning using PyTorch
syft.js	PyTorch and PySyft - Javascript frameworks enabling to run visual trainings in browser

Federated Learning and Differential Privacy (FL&DP) framework from Sherpa.AI	Simple, open-source FL framework integrating TensorFlow for deep learning SciKit-learn for linear models including privacy mechanisms
NVIDIA Clara Train SDK	Proprietary solution with FL support added from version 2.0. It uses TensorFlow and supports AutoML.

The FL frameworks for IoT are discussed in [AI-12].

3.2.6.2. Explainable AI

Explainable AI (XAI) is understood as a set of tools and techniques to help to understand and interpret predictions made by ML models. It has become crucial for understanding how an AI model reaches decisions and for identifying possible sources of errors. There are two approaches to achieve explainability: (i) build a transparent ML model, (ii) use black-box model and apply post-hoc technique to explain its behaviour. The former is a current and challenging research topic. For the latter different techniques can be used, including model-specific or model-agnostic, local or global, e.g. data visualization, decision tree, logistic regression model, neural network model, SHAP (SHapley Additive exPlanations).

The National Institute of Standards (NIST), part of the U.S. Department of Commerce, defines four principles of explainable artificial intelligence [AI-13]:

- An AI system should supply “evidence, support, or reasoning for each output.”
- An AI system should provide explanations that its users can understand.
- Explanation accuracy. An explanation should accurately reflect the process the AI system used to arrive at the output.
- Knowledge limits. An AI system should operate only under the conditions it was designed for and not provide output when it lacks sufficient confidence in the result.

NIST defines 5 types of explanation:

- Inform the subject of an algorithm.
- Build societal trust in an AI system.
- Satisfy compliance or regulatory requirements.
- Assist with further system development.
- Benefit the algorithm’s owner.

XAI can be divided into three categories that can be addressed separately:

- Explainable data
- Explainable predictions
- Explainable algorithms

One can also distinguish the following types of XAI [AI-17]: interpretable AI (user cannot only see, but also study and understand how inputs are mathematically mapped to outputs), transparent AI (user can see how AI operates using, e.g. summaries, visualizations, descriptions) and interactable AI (users can interact with the machine learning model to understand why it made a specific decision).

Table 2. Explainable AI tools

Tools for explainable AI	
Activation Atlases	Google collaborates with OpenAI to develop this technique to visualize the interaction between neural networks. It monitors the way neural networks expand their horizon with information and various layers.

AIX360	IBM framework to enable the interpretability and explainability of various datasets in a machine learning model. A Python package that includes comprehensive algorithms that monitor various dimensions of explanations and their proxy explainability metrics.
Alibi	An open source Python library aimed at ML model inspection and interpretation. It focuses on providing the code needed to produce explanations for black-box algorithms.
DeepLIFT	A comparative technique for activation of each neuron to its “reference activation”.
InterpretML	Microsoft toolkit aimed at improving explainability.
LIME	Local Interpretable Model-Agnostic Explanations is a technique developed by researchers from the University of Washington. It helps attain a higher level of transparency within an algorithm.
Shapley	SHAP (Shapley Additive Explanations) is a method to explain individual predictions, based on the game theoretically optimal Shapley values.
Skater	It provides model interpretation for all types of models. It also helps to develop an understandable machine learning system.
Rulex Explainable AI	Rulex is a company that develops predictive models for first-order conditional logic rules.
What-If Tool (WIT)	Framework by TensorFlow that visually represents datasets and provides comprehensive results.

Note that the above described tools were not designed for distributed AI or to function in edge-cloud continuum based solutions. The issue of how to address explainability in edge computing environments is an emerging area of study [AI-14].

3.2.6.3. Frugal AI

The current challenge is to develop new AI methods that are able to make use of less training data than current state-of-the-art deep learning algorithms while maintaining similar performance. Noteworthy is that data is crucial to provide an effective and efficient ML-based solution. Here, so-called frugal artificial intelligence systems that require less data and less computing power to build them, may be a solution. One can distinguish three options of frugality [AI-15]: input frugality (related to cost of data, may result in less training data or features), learning process frugality (related to cost of computational and memory resources) and model frugality (related to costs of storing and using ML model). In [AI-15] the authors outline the concept of frugal AI and experiment with supervised learning (classification) using data from smartwatches. A new framework for the analysis of machine learning algorithms in terms of their frugality, i.e., of how proficient they are at delivering accurate predictions when working with limited resources. A novel evaluation measure was introduced - the frugality score, which trades off predictive accuracy for resource consumption and can be adjusted to the resources available to a learning algorithm.

It is difficult to train a model from a relatively small amount of data, or even from a single instance (one shot learning). However, there are methods to overcome or bypass this lack of data. Moreover, AI models trained with smaller datasets can reduce compute resource requirements, storage infrastructure, data processing costs, energy costs.

Note that lack of data can refer to data itself or to labels (used in supervised learning) [AI-16]. Labeling data very often cannot be automated and require human involvement. Different techniques can be applied depending on the situation: (i) low data, low label - transfer learning, domain randomization, synthetic data, (ii) low data,

no labels - synthetic data, (iii) high data, low labels - self-supervision, self-distillation, synthetic data, (iv) high data, no labels - self-supervision, self-distillation.

3.2.6.3.1. Transfer learning

This technique is based on using knowledge gained while solving one problem to a different but related problem, e.g. reusing an existing pre-trained AI model that has learnt from a sufficient dataset that is similar to the missing data [AI-20, AI-21, AI-22]. It is a popular approach to deep learning problems where pre-trained computer vision and natural language processing models are used to save resources required to develop neural network projects. Depending on the task domain and the amount of labeled and/or unlabeled data available, transfer learning falls into three main categories:

- **unsupervised transfer learning** - a model is trained on a source dataset, and then used to learn a target task on a different dataset. It can be used when there is no labeled data available for the target task. When model is trained on multiple datasets its generalizability can be improved.
- **inductive transfer learning** – a model is learnt on a source dataset and then is applied to the target dataset. It is used when the source and target datasets are very different in size or structure.
- **transductive transfer learning** – is applied when there is a large dataset that should not be revealed, instead a small subset of data is provided for learning. One of the techniques that can be used is leave-one-out cross-validation.

Transfer learning is related to problem of multi-tasks learning and concepts drift. Multi-task learning is an ML approach in which multiple tasks (with some level of correlation) are learnt simultaneously, i.e. rather than training independent models for a group of tasks there is a single model for all of the tasks. Often all of the available data across the different tasks are used together to provide generalized representations of the data that can be used in multiple contexts. Concept drift is when model's predicted target variable or its statistical properties change over time (contrary to data drift when dataset changes over time).

The following subsections outline types of transfer learning that can be used in different scenarios.

3.2.6.3.1.1.Zero-shot learning

Zero-shot learning [AI-18] is a method where a pre-trained model is used to evaluate test data of classes that have not been used during training, i.e. ability to complete a task without having received training examples for it. Zero-shot methods work by associating observed and non-observed classes usually through some form of auxiliary information, which encodes observable distinguishing properties of objects. An example of such approach is zero-shot translation in the Neural Translation model (GNMT) by Google that offers cross-lingual translations. Translation between two discreet languages is done with a pivot language. For instance, if translation needs to be done from Norwegian to Japanese, first Norwegian is transferred to English and then from English to Japanese. The translation uses data to learn the translation techniques for language pairs.

3.2.6.3.1.2.One-shot learning

One-shot learning [AI-26] is a method of learning information about object categories from one training example by treating classification problem more like difference-evaluation problem. Most popular usage area is computer vision e.g. facial recognition, documents check [AI-25]. One-shot deep learning model takes two images (e.g., the image from the document and the image of the person looking at the camera) and returns a value representing the similarity between the two images based on which it can be indicated if this is the same person with respect to a predefined threshold.

3.2.6.3.1.3.Few-shot learning

Few-shot learning [AI-19, AI-23] is an example of meta-learning, where training is done on several related tasks, so that it can generalize well to unseen (but related) tasks with just a few examples. The most common application areas of few-shot learning are: computer vision, natural language processing (parsing, translation, sentence completion), audio processing (voice cloning, conversations). Scenarios addressed include models that imitate human cognition (learning from a few examples), models that can be generalized across similar tasks, models that should recognize rare cases.

3.2.6.3.2. Active learning

Active learning is a method used in situations when data is available but without labels and labelling is expensive [AI-27]. This semi-supervised learning introduces “an oracle” into the process. The algorithm formulates queries i.e. chooses data to be labelled by the oracle. The goal is to find optimal queries with respect to information gain, i.e. to select data which should be labelled in order to have the highest impact to training a supervised model. Two types of sampling can be distinguished: stream-based (unlabelled data is continuously fed to an active learning system, the learner decides whether to send data to a human oracle based on a learning strategy) and pool-based (the data samples are chosen from a pool of unlabelled data based on the informative value scores and sent for manual labeling). Strategies for subsampling include: committee based strategies, probability-based strategies and large-margin based strategies. Popular frameworks for active learning are: modAL, alpacaTag and libact.

3.2.6.3.3. Hybrid AI

Hybrid AI [AI-24] means combining different tools/algorithms (e.g. non-symbolic AI with symbolic AI or human intelligence) to address a problem from different angles, using different models in order to deliver optimal output and require less data for training.

3.2.6.3.4. Data generation and data augmentation

The goal in using this technique is to generate artificial data but as close as possible to data coming from a real environment using a simulation environment. Then this data can be used for training. If the data cannot be generated “from scratch” then data augmentation techniques can be used to generate new data from existing data. Data augmentation techniques for computer vision include: adding noise, cropping, flipping, translation, scaling, rotation, brightness, saturation etc. Augmentation techniques in natural language models include: synonym replacement, text substitution, random insertion/swap/deletion, back translation, text generation etc.

3.2.7. Security, integrity, trust, privacy and policy enforcement in the computing continuum

The massive growth of IoT devices and in extend to the huge amount of data traffic have created additional issues on the bandwidth and resources of the centralized cloud computing paradigm. The recent advancements in the computer continuum have contribute on tackling this issue by employing the edge computing strategy. Even though, this strategy improves the QoS, it has introduced additional issues in data security, privacy, and trust. Furthermore, the last couple of years both academia and industry focus on the enhancement of security and privacy aspects in the computer continuum as well as on the evolvement of trust mechanisms between different components in Edge-Cloud (EC) architectures. Following, are discussed the major challenges and prominent solutions in terms of security, privacy, and trust in EC architectures.

3.2.7.1. Edge-Cloud Security

Challenges-Issues

The main issue regarding the security of the EC is the protection of the data and the components that constitute the EC infrastructure. Malicious actions in the EC can be encountered during the three main processes, namely communication, computation, and storage. The literature has identified and studied several attacks that can be occurred at different levels and layers (e.g., EC devices, communication and EC servers/nodes, and cloud servers). The main attacks/challenges that have been identified are:

1. Malicious hardware/software injection: Adversaries can add unauthorized software and hardware components to the communication or node levels of the infrastructure. The malicious injections will aim to exploit service providers to perform malicious actions on their behalf, such as bypassing authentication, stealing data, exposing database integrity or reporting false data. As one can understand, this type of attack can have serious consequences on the EC infrastructure compromising the whole infrastructure [SCC-1]

2. Denial of Service: Adversaries flooding the network with counterfeit messages to exhaust communication, computing, and/or storage resources. This will have as a result authorized users to not be able to use the EC services.
3. Eavesdropping/Sniffing: Attackers secretly monitor communication links to obtain access or control information of the EC nodes, such as node's configuration, identifiers, and passwords. Acquiring this information unauthorized users can obtain access on the EC infrastructure.
4. Security threats from/on IoT devices: Mobile botnets, IoT malwares, and ransomwares are on the rise affecting both edge users and applications leading to application freeze up or data leakage.
5. Causative attacks against machine learning models: Machine learning (ML) is heavily used on EC applications and are often target of causative attacks, namely attacks that manipulate or inject misleading examples on the training dataset. Causative attacks on ML might affect the performance of the ML models and based on the model's task resulting in various issues (e.g., insufficient security, unexpected actions, etc.).
6. Policy violation: Malicious actions that violate the existing policies of an EC infrastructure.

Solutions

1. Intrusion Detection System (IDS): IDSs are essential security solutions that play a key role for the protection of EC continuum against malicious actions.
2. Securing firmware updates: IoT devices should always be up to date to the latest firmware updates. Therefore, the firmware updates should be performed in an automatic manner.
3. Access control: Authentication, identity management, and access control are of utmost importance in EC applications to maintain the security of applications and protect them from unauthorized access. Access control should answer three questions [SCC-2]: *Who should have access? What should access? For how long should have access?*
4. Policy-based mechanisms: Employing a set of policies to manage the systems in an EC ecosystem. The policy-based mechanisms can be used to detect violation of policies by assuring that standard rules are applied and are not breached [SCC-3].

3.2.7.2. Edge-Cloud Privacy

Challenges-Issues

Ensuring the privacy of the EC infrastructure is a challenging task to be achieved. Several issues have been identified in the literature related with maintaining the privacy in EC. The main challenges as well as novel solutions that enhance the protection of privacy are discussed below.

1. Data privacy: Huge amounts of data are processed and stored within the EC ecosystem, which have been acquired from applications or users' devices. Despite the security and trustworthiness level that has been applied in the EC, there are always threats that might compromise its operation. Thus, maintaining the privacy of the data is also crucial and challenging.
2. Privacy leakage: For some operations EC devices might need to obtain personal information from the data, regardless of whether this information is sensitive it must belong to information owners. However, this information could be transmitted with other users or entities within a network without owner's permission. This could make them vulnerable to attackers during the data transmission.
3. Location data leakage: Several applications in order to be fully functional utilized the device's location (e.g., in smart home environments). Location data are sensitive, and its leakage violates the users' privacy.

Solutions

1. Cryptographic primitives: In 2021, only 24% of IoT devices encrypt the data before transmission¹, namely the remaining 76% transmit data unencrypted. Thus, provable secure cryptographic schemes should be deployed to encrypt the data both when stored and transmitted within EC continuum. In this

¹ <https://www.venafi.com/blog/cyber-attacks-iot-devices-are-growing-alarming-rates-encryption-digest-64>

way, an extra level of protection will be employed on the data enhancing the total data privacy of the systems and reducing the possibility of privacy leakage.

2. **Anonymization:** Data anonymization methods are often an essential solution to preserve data privacy and due to the fact that it is an active research field for almost two decades, several innovative solutions have been proposed over the years. The main purpose of anonymization is to hide information related to Personally Identifiable Information (PII) (e.g., names, credit card details, mobile numbers, etc.) [SCC-4]. State-of-the-art methods for maintaining anonymity in EC are k-anonymity, l-diversity, T-closeness, differential privacy, and hashing functions [SCC-5].
3. **Decentralization:** Distributing the sensitive information through various EC nodes, hence no node will have full knowledge of the information.
4. **Secure Data Aggregation:** A privacy-preserving data compression strategy that is based on homomorphic encryption to encrypt data and then send it to the EC nodes. The EC nodes aggregate the data to calculate the multiplication of individual data and send the aggregated results to cloud servers [SCC-6].

3.2.7.3. Edge-Cloud Trust

Challenges-Issues

Trust is a critical issue in the EC environment and trust management is a key element to draw the users' attention in EC applications. Trust is the combination of security, privacy, and availability. Hence, in order to establish trust in a system, the security, privacy, and availability should be solid. However, maintaining trust in a complex environment such as EC is not an easy task and comes with several challenges [SCC-7]. Some of these challenges are listed below:

1. **Identification of the level of trust:** A concern regarding the trustworthiness of EC ecosystems is how much trust the various entities of the EC they can put on each other when they exchange data and work with each other. This trust level should be visible and transparent in the most possible extend.
2. **Defamation (bad-mouth attack):** A common attack that targets trust management systems that includes "bad" nodes, which recommend incorrect values about the neighboring nodes of the network aiming to reduce their reputation. This attack may have a severe effect on the EC environment depending on the number of "bad" nodes.
3. **Handoff attack:** Migrating a device from one place to another in a network looking for new EC nodes to sink. It is often used to perform malicious activities, such as reduce network performance and [SCC-2] consume network resources.
4. **Collision attack:** A group of malicious nodes collaborate to influence the trust level of EC nodes.

Solutions

1. **Trust evaluation:** Is a common method for enhancing the degree of trust in EC. Particularly, Gao et al. [SCC-8] introduced a multidimensional trust evaluation method to solve the trust evaluation problem on edge devices in IoT environments. In [SCC-9] the authors proposed a trust evaluation method based on crowdsourcing and hierarchical trust management for trust evaluation in cyber-physical and cloud computing systems. A trust computation framework proposed by [SSC-10] that leverages black/white-lists to select trusted communication parties.
2. **Holistic Trust Management:** Hybrid authentication and authorization combining self-sovereign identities, distributed identifiers, verifiable credentials, as well as FIDO, TEE, and hyperledger fabric for storing trust scores.

The following table gathers the security, privacy, and trust challenges in the EC continuum that mentioned in the previous sections along with the solutions that have been developed to tackle these challenges.

Table 3. Security, Privacy, Trust Challenges and Solutions

Category	Challenge	Solution
Security	Malicious hardware/software injection	Access control, IDS

	Denial of Service:	IDS
	Eavesdropping/Sniffing	Access control
	Security threats from/on IoT devices	Securing firmware updates
	Causative attacks against machine learning models	Access control, IDS
	Policy violation	Policy-based mechanisms
Edge-Cloud Privacy	Data privacy	Cryptographic primitives, Decentralization
	Privacy leakage	Decentralization, Secure Data Aggregation
	Location data leakage	Anonymization
Edge-Cloud Trust	Identification of the level of trust	Trust evaluation
	Defamation (bad-mouth attack):	Trust evaluation, Trust management
	Handoff attack	Trust evaluation, Trust management
	Collision attack	Trust evaluation, Trust management

In order aerOS to tackle the aforementioned challenges a holistic solution will be developed that will be based on holistic trust management. Particularly, authentication, authorization, and trust management methods will be combined including technologies such as self-sovereign identities, distributed identifiers, verifiable credentials, TEE, and blockchain (e.g., hyperledger fabric).

3.2.8. From DevOps to DevSecOps to DevPrivSecOps

The project will follow the DevPrivSecOps approach as a continuous model, and thus, will contribute to develop security, privacy and operation in systems that require continuous privacy and security. This is an evolution from the DevOps methodology, that is a well-known industry standard for software development in a continuous, fluid and agile way; first to DevSecOps approach with the aim of including security concerns and controls in all the phases of the SW development cycle, so that cybersecurity can be considered and included by design; and then to DevPrivSecOps with the purpose of increasing the privacy knowledge and partnership of developers, testers, operations staff and security experts.

The next figure shows a high-level overview of the evolution from DevOps to DevSecOps to DevPrivSecOps.

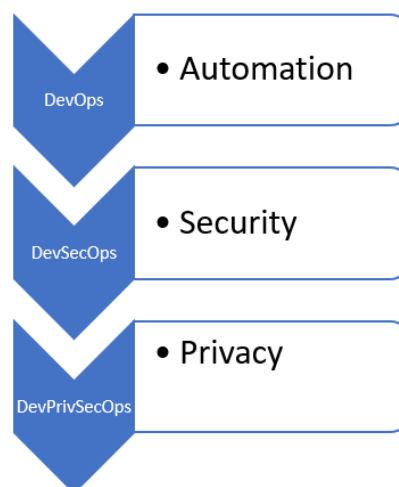


Figure 41. Evolution from DevOps to DevSecOps to DevPrivSecOps

This section will highlight the main characteristics of each methodology and the motivation to go a step further to include aspects of security and privacy in the process.

3.2.8.1. DevOps

Historically, the lack of cooperation among the development and operations teams in software production often resulted in facing a lot of challenges along the software development lifecycle. Hence, the plan of deploying so many changes at once leads to very hard forensics processing on identifying what, where and why are located those bugs that crashes the new release available.

This is where DevOps came into play. The term coined by Patrick Debois, in October 2009 [DPSO-1] is about fast, flexible development and provisioning of business processes, which by efficiently integrating development, delivery, and operations, facilitates a lean, fluid connection of these traditionally separated silos [DPSO-2]. The most consolidated definition of DevOps [DPSO-3] is: "DevOps is a collaborative and multidisciplinary effort within an organization to automate continuous delivery of new software versions, while guaranteeing their correctness and reliability".

DevOps integrates the two worlds of development and operations, using automated development, deployment, and infrastructure monitoring. It is an organizational shift in which, instead of distributed siloed groups performing functions separately, cross-functional teams work on continuous operational feature deliveries. This approach helps to deliver value faster and continuously, reducing problems due to miscommunication between team members, and accelerating problem resolution.

There are various phases in the DevOps lifecycle. The DevOps lifecycle refers to a continuous software development process that uses DevOps best practices throughout the lifecycle of the software. It is often presented in a continuous loop. Although there are several approaches aiming to identify which are the different DevOps stages or phases, those that are most frequently adopted in DevOps culture includes eight phases: Plan, Code, Build, Test, Release, Deploy, Operate, Monitor, as presented in the following figure.

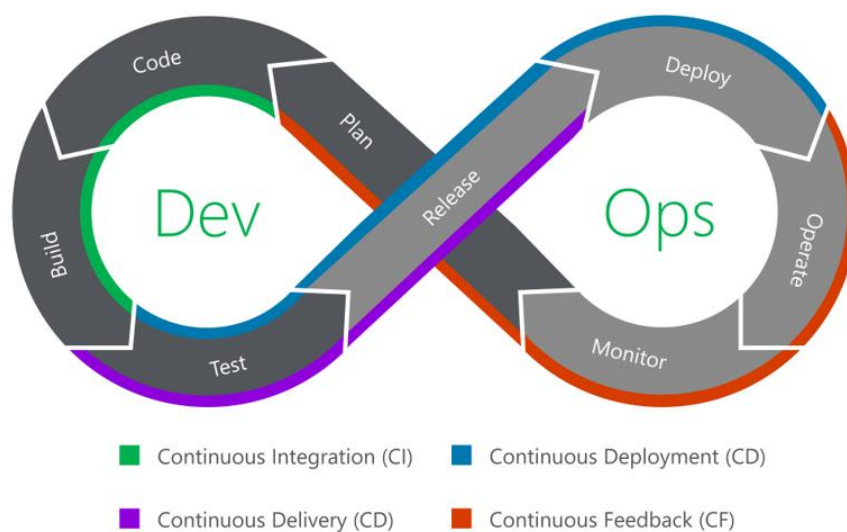


Figure 42. DevOps workflow

A short description of phase [DPSO-4] is described next:

- **Plan:** The Plan stage covers everything that happens before the developers start writing code, and it is mainly relate with the Product/Project Manager role. Requirements and feedback are gathered from stakeholders and/or customers and used to build a product roadmap to guide future development.
- **Code:** This is the phase where the developments start. In addition to the standard toolkit of a software developer, the DevOps team has a set of plugins installed in their development environments to aid the development process, including consistent code-styling and avoiding common security flaws. Resulting in developers good coding practice and in fewer failed builds.

- **Build:** Once a developer has finalized a task, the resulting code is committed to a shared code repository, typically through a pull request. Another developer then reviews these changes and once there are no issues, the pull-request is approved. Simultaneously, the pull request triggers an automated process, which builds the codebase and runs a series of tests to identify any regressions. If the build fails, or any of the tests fail, the pull-request fails, and the developer is notified to resolve the issue.
- **Test:** Once a build succeeds, it is automatically deployed to a staging environment for deeper, out-of-band testing. Once the application is deployed to the test environment, a series of manual and automated tests are performed.
- **Release:** The Release phase is a milestone in a DevOps pipeline, as it is the point where a build is ready for deployment into the production environment. By this stage, each code change has passed a series of manual and automated tests, and the operations team can be confident that breaking issues and regressions are unlikely.
- **Deploy:** This stage is when a build is released into production. The new environment is built, and it sits alongside the existing production environment. When the new environment is ready, the hosting service points all new requests to the new environment. If at any point, an issue is found with the new build, it is just necessary to tell the hosting service to point requests back to the old environment.
- **Operate:** The new release is now live and being used by the customers. In this stage, the operations team should make sure that everything is running smoothly. It is recommended to build a way for the customers/stakeholders to provide feedback on their service.
- **Monitor:** The final phase of the DevOps cycle is to monitor the environment, sustained by the customer feedback, by collecting data and providing analytics on customer behavior. All this information is fed back to the Product Manager and the development team to close the loop on the DevOps process. This should be considered as a DevOps continuous process.

3.2.8.2. DevSecOps

In the past, the role of security was isolated to a specific team in the final stage of development, but those days are over. Now, in the collaborative framework of DevOps, security is a shared responsibility integrated from end to end. Security is so important that it led to coin the term “DevSecOps” to emphasize the need to build a security foundation into DevOps initiatives.

DevSecOps [DPSO-5] means thinking about application and infrastructure security from the beginning and also embedding DevOps with security controls providing continuous security assurance. DevSecOps is a natural extension of DevOps to include security-by-design and continuous security testing by automating some security controls in the DevOps workflow. Next figure presents how DevSecOps embeds security controls across the DevOps lifecycle phases.

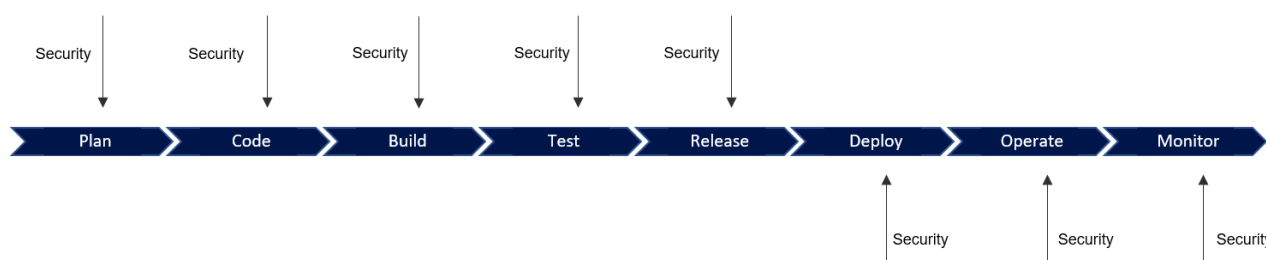


Figure 43. security Controls in the DevSecOps workflow

The core concept of DevSecOps is that everyone is responsible for security. Management must take into consideration when defining requirements and developing schedules. Developers must incorporate it into every facet of code and specifications. Security must be tested by QA professionals in addition to functionality. Finally, operations teams must monitor software behaviour and respond quickly to problems. Therefore, security awareness must be incorporated into each stage (Plan, Code, Build, Test, Release, Deploy, Operate, Monitor) [DPSO-6].

- **Plan:** The planning phases involves collaboration, discussion, review, and a strategy for security analysis. Teams must conduct a security analysis and develop a schedule for security testing that specifies where, when, and how it will carry it out.
- **Code:** Developers can produce better secure code using DevSecOps technologies during the code phase. Code reviews, static code analysis, and pre-commit hooks are important code-phase security procedures. Every commit and merges automatically should start a security test or review when security technologies are directly integrated into developer's workflow.
- **Build:** In this step the primary objective of DevSecOps build tools is automated security analysis of the build output artifact. Static application software testing (SAST), unit testing, and software component analysis are crucial security procedures. Tools can implement into an existing CI/CD pipeline to automate these tests.
- **Test:** Dynamic application security testing (DAST) tools are used throughout the testing process to detect application flows such as authorization, user authentication, endpoints connected to APIs and SQL injection.
- **Release:** This stage focuses on protecting the runtime environment architecture by reviewing environment configuration values, including user access control, network firewall access, and personal data management. One of the main concerns of the release stage is the principle of least privilege (PoLP), it signifies that each program, process, and user need the minimum access to carry out its task and combines checking access tokens and API keys to limit access for the owners.
- **Deploy:** The security problems that only affect the live production system should be addressed during deployment. It is essential to carefully examine any configuration variations between the current production environment and the initial staging and development settings. The deploy stage is a good time for runtime verification tools to gather data from an active system to assess if it functions as intended.
- **Operation:** Operation teams should monitor vulnerabilities frequently. DevSecOps should use appropriate tool to protect the organization infrastructure from cyber threats.
- **Monitor:** A breach can be avoided if security is constantly being monitored for anomalies. It is essential to deploy a robust continuous monitoring tool that operates in real-time to maintain track of system performance and detect any exploits at an early stage.

3.2.8.3. DevPrivSecOps

DevPrivSecOps is an evolution from DevSecOps with the purpose of increasing the privacy knowledge and partnership of developers, testers, operations staff and security experts, as shown in the following figure.

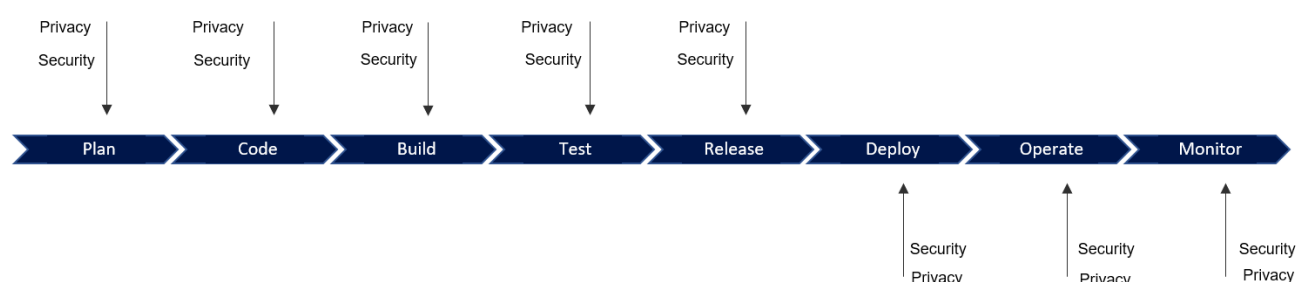


Figure 44. Privacy considerations in the DevPrivSecOps workflow

Again, DevPrivSecOps is about introducing automated privacy controls in the DevOps workflow. A key point will be to introduce privacy techniques in the SW design phase, in order to take into account mechanisms such as anonymization of data and data separation and to take privacy concerns into consideration when storing this data.

DevPrivSecOps methodology will be further developed and detailed in “D2.4 DevPrivSecOps Methodology specification V1” and “D2.5 DevPrivSecOps Methodology specification V2”, due to M9 and M21 of the project respectively.

3.2.9. Distributed multiplane analytics

Distributed analytics spreads data analysis workloads over multiple nodes in a cluster of servers, rather than asking a single node to tackle a big problem. The same algorithms run across each of the nodes, processing a subset of the data. Traditionally analytics involved the collection, transportation and processing of large data in the backend. Large data sets were preferable as they often produced more insights than smaller ones. Therefore, research focused on the scaling challenges to keep up with growing data sets, these led to properties such as Map Reduce, Hadoop and Hive. Applying this form of analytics to the networking domain is especially challenging as the data sets grow exponentially as the networks become larger. The effectiveness of analytics depends on the networks ability to generate telemetry data which is limited by available resources. Also monitoring interfaces often expose a polling mechanism which is available periodically and provides a sampled version of monitoring data. This may be problematic for time sensitive and complex diagnostic scenarios. “Big Data” techniques are limited by the availability of useful data in the network domain. This coupled with the inherent wastefulness of many Big Data applications, where large amounts of generated data are effectively thrown away while in hindsight other data may have been more valuable. Such as real-time network monitoring data focused on outliers and anomalies or summarizations and aggregations of particular areas of the network. Instead of collecting data from many locations to be made centrally available for processing, distributed analytics provides this processing at the source. This is achieved through a software function embedded in the network device allowing for more control over what data is generated. Data sources can be set up and adjusted to generate exactly the data needed to support the analytics task. Instead of large volumes of raw data, devices export small volumes of condensed information, or Smart Data [DMA-1].

The proliferation of data in modern networks has led to increased analytics from both the management and business perspective. Data-driven functions and services rely on insights generated using large data sets comprised of information from a range of components and users. Most current approaches to process this data utilise centralised, cloud-based storage. Models can be trained using large diverse data sets on resources available through the public cloud. This allows for straightforward analyses of the datasets but also incurs several disadvantages:

1. Changes in regulation may affect the risk and cost of centralised storage.
2. The processing and/or merging of multiple independent datasets is complicated when working with algorithms initially designed for centrally stored data.
3. Data-sharing is restricted due to uncertain commercial risks.
4. Data collection is becoming more restricted

Alternatively, distributed data analytics moves the code and models for training away from the centralised cloud and closer to the location where data is collected. This approach has been enabled through the increased processing power and memory capacity of devices at the edge of the network. Concerns around privacy and security has also motivated the distributed approach as they carry less risk when compared to centralised storage. The distributed approach is also viewed as being more energy efficient, reducing the movement of large data around the network frees up resources for other services improving performance for current users or allows them to be spun down in during quite periods [DMA-2].

Distributed analytics in the IoT domain is a continued research topic both in terms of tooling for straightforward implementation and deployment efforts and distributing computational workloads around the system. In [DMA-3] the authors design and develop a configurable engine for distributed data analytics for IIoT applications. The engine utilises state of the art data streaming middleware platforms and updates with new digital models reducing the effort needed to implement and deploy distributed data analytics in IIoT environments. In [DMA-4] the authors present fog-specific decomposition of multivariate linear regression and apply the decomposition method to the analytics model to run in a distributed manner in the fog-enabled IoT deployments. The approach avoids sending raw data to the cloud and offers balanced computation in the infrastructure. In [DMA-5] the authors acknowledge the challenges of distributed deployment of DNN models onto resource-constrained fog

nodes with low latency in IoT domain. Model compression techniques and horizontal model partition techniques are identified as existing solutions with limitations. Alternatively, the authors propose an integrated Efficient Distributed Deep Learning (EDDL) framework to addresses previous limitations through a Balanced Incomplete Block Design (BIBD), joint horizontal and vertical model partition and multi-task and ensemble learning techniques.

The research community has put large efforts into bringing concepts from distributed analytics to “Big Data”. This is pursued to produce faster and more efficient results to the “Big Data” approach. In [DMA-6] the authors introduce a distributed and self-organizing algorithm to build a management system for big data analytics in the healthcare domain. Local autonomous operations performed by hosts in the distributed system feed resource discovery operations making them faster and more efficient. Also, in [DMA-7] the authors acknowledge the cost, time and scalability issues of “Big Data” and motivate the need for alternative approaches such as data-less “Big Data” where analytics is performed by employing learned models of data and queries instead of accessing any raw data. A distributed approach to “Big Data” is a key consideration moving forward. In [DMA-8] the authors provide a current state of the art for Big Data management through a number of topics from advanced “Big Data”, Privacy Preserving “Big Data” and imprecise “Big Data”. All of these topics look at “Big Data” in distributed environments. The authors have also provided considerations for driving future research efforts in the field. In [DMA-9] the authors acknowledge that “Big Data” is becoming more stream oriented and data is processed as it arrives by distributed and low-latency computational frameworks. The authors provide a comparative study of distributed data stream processing and analytics frameworks. The authors also present a critical review of representative open source and commercial distributed data stream processing frameworks.

The visualisation of distributed analytics is also a research consideration with emphasis on simplifying the process for users. In [DMA-10] the authors propose a visual analytics framework that addresses the complex user interactions required through a command-line interface to run analyses in distributed data analysis systems. The visual analytics framework facilitates the user to manage access to the distributed servers and provides a number of analysis and visualisation functions to the user.

Table 4. Libraries and Tools for Distributed AI

Libraries and Tools for Distributed AI	
Dask	Dask offers a distributed framework as a task-based environment to compute resource using dynamic task scheduler. The central dask scheduler in dask coordinates the actions of several dask worker, which processes multiple machines and caters to concurrent requests of several clients.
Dataiku	Dataiku data science platform which allows analysts and data scientists to build predictive applications more efficiently and deploy them into a production environment. It supports a range of features and applications.
RAY	Ray is an open-source unified compute framework that makes it easy to scale AI and Python workloads — from reinforcement learning to deep learning to tuning, and model serving.
pandas	pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.
scikit-learn	Scikit-learn is an open-source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data pre-processing, model selection, model evaluation, and many other utilities.
NumPy	NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

Practically, all existing implementations of algorithms operate with the training set entirely in main memory. If the computational complexity of the algorithm exceeds the main memory then the algorithm will not scale well, will not be able to process the whole training data set or will be unfeasible to run due to time or memory

restrictions. Thus, in order to handle “very large” data sets, a new and active research field emerges, large-scale learning techniques such as dask with Dataiku, RAY, Federated Learning (Edge AI) and Efficient Distributed Deep Learning (EDDL) often increases the accuracy achieved. It intends to develop efficient and scalable algorithms with regard to accuracy and to requirements of computation (memory, time and communication needs) [DMA-11].

3.3. Surrounding ecosystem

As organizations around the world embrace digital transformation more and more data from Internet of Things (IoT) devices, smart sensors, and other devices are being generated on the edge of the organizations’ networks. Those data are collected, stored, and processed across clouds, edges, data centres, and colocations, and thus each organization must re-examine the ability of its existing technology to meet the demands of the data growth, edge expansion, IoT, and distributed workforces. Additionally, new applications create a growing need for real-time data-driven decision making, especially at the edge, that could be negatively affected by the quick data growth.

While many data still reside on premises, other types of data are collected, processed, and managed at the edge – outside of traditional data centers or public clouds – and are expected to grow significantly in the near future, managing workstreams across these remote sites, in addition to ones on-premises is a challenging task.

Tackling the aforementioned challenges, an edge-to-cloud approach is designed to bring the cloud experience to all of an organization’s apps and data, regardless of where they may reside. Following this trend, EU Data Strategy sets the edge-to-cloud hybrid paradigm as a strategic technology towards European leadership in the digital space.

aerOS and other DATA-01-05 cluster projects will contribute to the desired outcomes by strengthening Europe’s supply and value chains in cloud-to-edge computing. More specifically, aerOS developments and impact are significant in order to ensure EU market leadership in distributed and decentralized data processing since aerOS will break the current circle of small-scale ad-hoc industrial edge implementation solutions and transition industrial stakeholders through a virtuous pathway of industrial IoT-edge economies of scale and open multi-sectorial solution provisioning.

3.3.1. Industrial approach to edge-cloud continuum in Industry (I4.0 and I5.0)

3.3.1.1. From Industry 3.0 to 5.0

The advances made in manufacturing technologies, industrial processes and other scientific areas such as physics, electronics and computing has led to several industrial revolutions throughout history, aiming at improving throughputs, reducing downtimes and lowering costs. It all started with the first industrial revolution in the form of mechanization and the steam engine, helping to accelerate the economy. The second industrial revolution came with the creation of the internal combustion engine, new methods of communications (telegraph and telephone) and the invention of automobile. Finally, in the second half of the 20th century the third and the most recent industrial revolution took place: The third Industrial Revolution.

The third industrial revolution (**Industry 3.0**) comprised an important advance related with the utilisation of field-level computers and automation ruled the industrial scene. I3.0 is considered to begin, alongside computer era, in 1950’s. During this era, automation tools (Programmable Logic Controller, robots) and technologies were introduced in manufacturing process, enabling the automation of tasks that were previously carried out by humans. The most important advances that took place in the third industrial revolution can be seen in the Figure 45.

Factory automation

- The introduction of new automation technologies (e.g robotic arms, automated assembly lines...) enabled an increased productivity, improved quality robustness, thus reducing direct human labor costs and errors.

Emergence of PLCs and microprocessors

- The development of advanced hardware such as Programmable Logic Controllers (PLCs) and microprocessors enabled a programmable industrial environment, leveraging a full automation, data acquisition and advanced control of complex industrial processes.

Development of industrial robotics

- The development of new technologies for advanced robotics allowed fully programmable automated lines which were capable of movement on three or more axes, enabling automation on previously manual applications such as welding, painting, packaging, labelling and product inspection/testing.

Development of supercomputers

- The latest advances in microprocessors allowed the development of high level performance computers, which enabled high data processing capabilities for new automation tools and robotics.

Business software

- In line with the previous technological advances, the development of computational technologies allowed the creation of advanced software that can exploit the new data generated in manufacturing processes and other business units in a company, e.g ERP, CRM, MRP...

Arrival of the internet

- The arrival of internet and World Wide Web is considered as the most important revolution in communication technologies. This allowed a very efficient way to share data and information between companies and remotely located machines, which enabled further advances in automation techniques, data sharing and processing.

Figure 45. Industry 3.0 most important technological advances.

The latest developments in advanced technologies such as Cyber Physical Production Systems (CPPS) and smart devices lead to the fourth industrial revolution (**Industry 4.0**) which is an initiative originated in 2011 from a project in the high-tech strategy of the German government. Industry 4.0 comprises rapid advances and changes in interconnectivity of processes and factories, as well as smart automation which is allowed due to the fast development of artificial intelligence and advanced robotics that integrate physical, digital and biological worlds.

The industry 4.0 relies on big-data technologies and Internet of Things (IoT), as the interconnectivity between complex production processes are expected to generate big data volumes within an advanced I4.0 facility which, at the same time, require connexion reliability, low latency and high computation performance. This is where edge-cloud continuum approach plays a vital role as Industry 4.0 enabler.

The most important advances in Industry 4.0 are shown in Figure 46.

Emergence of smart factory

- The smart factory concept comprises an ideal environment in which an industrial facility can be automatically operated without human intervention.

Data-driven automation

- The automation of processes in Industry 4.0 framework no longer depends exclusively on programmable controllers, but it acquires a higher dimension by taking decisions based on data inputs.

Machine-to-Machine communication

- As an industrial process must be interconnected, industrial devices are capable to communicate each other, taking automated decisions by data-driven automation.

Predictive capabilities

- Based on advanced data-analytics and new IA technologies, an intelligent industrial process is capable to predict accurately different scenarios.

Figure 46. Industry 4.0 most important technological advances.

Currently, a new concept that complements Industry 4.0 has been defined driven by the impact of the pandemics, which is considered the **Industry 5.0**. The term Industry 5.0 is not considered an industrial revolution itself, as it is a complement or a correction of the concept Industry 4.0.

The Industry 5.0 comes to solve one of the most intimidating facts of I4.0: a fully automated factory would not require the presence of human intervention for its successful operation. In contrast, Industry 5.0 brings back empowered humans to factories by using advanced technologies, as shown in Figure 47.

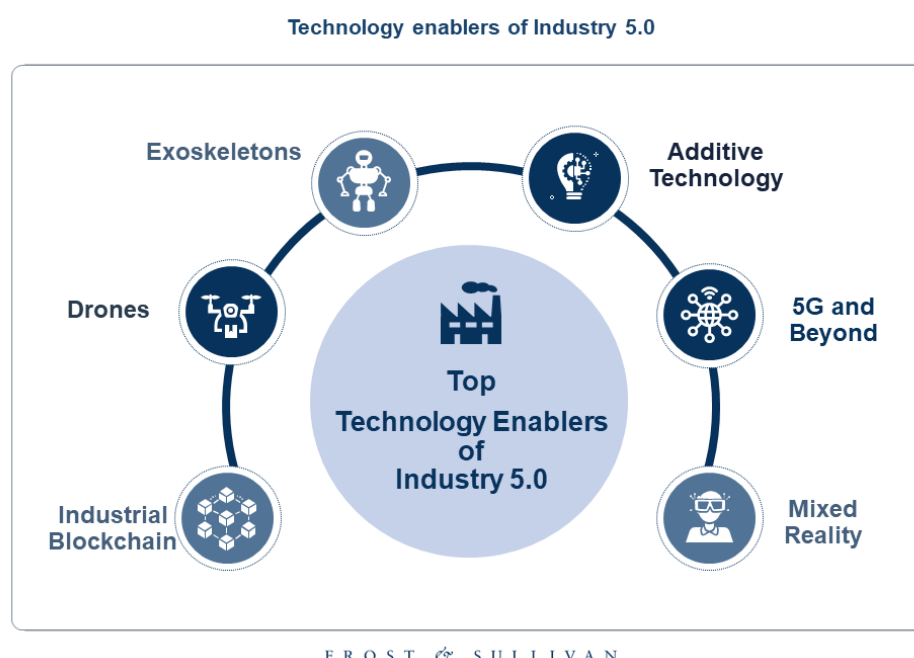


Figure 47. Technology Enablers of I5.0 [IECC-1]

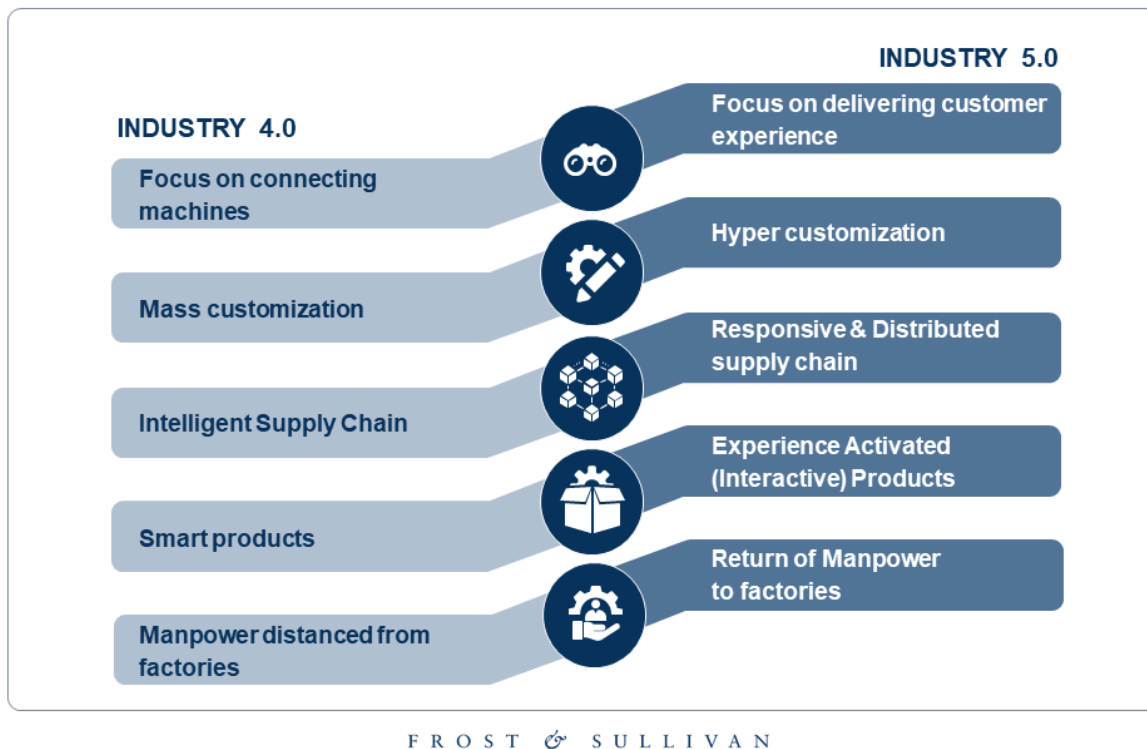
On the other hand, I5.0 brings new business concepts aided by advanced technology themes, focused on delivering tailored customer experiences by defining products, services and solutions individually. This will

drive market interests toward hyper customization, as each individual product will be unique to its intended customer and manufactured accordingly.

To achieve this, manufacturers will have robotized intelligent factories around the globe to manufacture the basic design of the product in bulk. The basic material will then be sent to local factories, where the final stages of the product will be completed using manual labour [IECC-2].

As it is shown in Figure 48, the main differences between I4.0 and I5.0 are related to customer experience, customization, a distributed supply chain, interactive products and the use of manpower in factories.

Highlights of Industry 5.0 compared to Industry 4.0



Fuente

especificada no válida.

Figure 48. Industry 5.0 compared to Industry 4.0 [IECC-1]

3.3.1.2. Reference Architectures for Edge-Powered I4.0

In this section, different existing initiatives regarding reference architectures for Industry 4.0 will be discussed. While currently exist reference architectures specifically designed for Industry 4.0 and edge computing, other kind of reference architectures regarding data spaces and zero-defect manufacturing will be also presented, as their functionalities are potentially applicable to aerOS project.

3.3.1.2.1. Reference Architecture Model for Industry 4.0

Currently, one of the most popular initiative regarding reference architectures is **The Reference Architecture Model for Industrie 4.0 (RAMI 4.0)**. It was first defined by the German Electrical and Electronic Manufacturer's Association (ZVEI), aiming to support Industry 4.0 initiatives as they are gaining broad acceptance throughout the world.

RAMI4.0 is a three-dimensional map showing how to approach the issue of Industrie 4.0 in a structured manner. RAMI 4.0 combines all elements and IT components in a layer and life cycle model, breaking down complex processes into easy to grasp packages, including data privacy and IT security. This gives companies a framework to approach the deployment of Industry 4.0.

RAMI 4.0 defines a service-oriented architecture (SOA) where application components provide services to the other components through a communication protocol over a network **Fuente especificada no válida..** The basic principles of SOA are independent of vendors, products, and technologies. The goal is to break down complex processes into easy-to-grasp packages, including data privacy and information technology (IT) security, as it is shown in Figure 49.

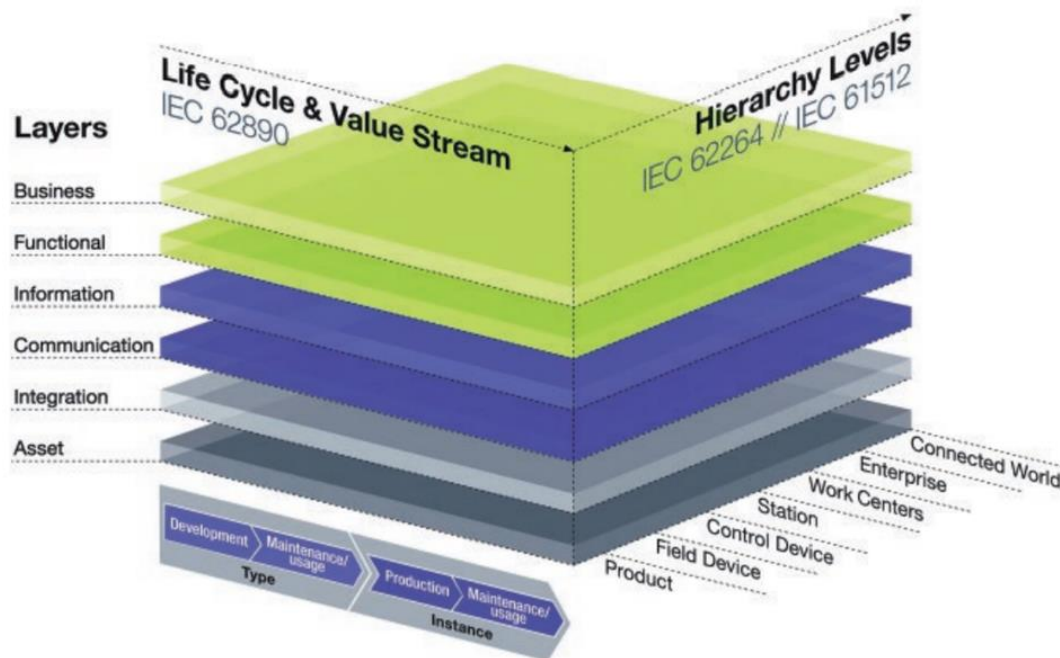


Figure 49. Reference Architecture Model for Industry 4.0 [IECC-3]

Each of the axis of RAMI 4.0 are further detailed in the following lines:

- **Axis 1: Hierarchy Levels**

There is a big difference between the Industrie 3.0 hierarchy levels and the Industrie 4.0 ones. While the Industrie 3.0 was a hardware-based structure, hierarchy-based communications and the product was isolated, the latter brings a new paradigm based on distributed functions throughout the network, communication among all participants and the product as a part of the network.

The right horizontal axis corresponds to hierarchy levels IEC 62264, the international standards series for enterprise IT and control systems. These hierarchy levels represent the different functionalities within factories or facilities. To represent the industry 4.0 environment, these functionalities have been expanded to include work pieces, labelled "Product," and the connection to the Internet of Things and services, labelled "Connected World."

- **Axis 2: Product Life Cycle:**

The left horizontal axis represents the life cycle of facilities and products, based on IEC 62890, Life-cycle management for systems and products, used in industrial-process measurement, control, and automation. Furthermore, a distinction is made between "types" and "instances": A "type" becomes an "instance" when design and prototyping have been completed and the actual product is being manufactured. The model also combines all elements and IT components in the layer and life-cycle model. This Axis is graphically detailed in Figure 50.

The Product: From the First Idea to the Scrapyard

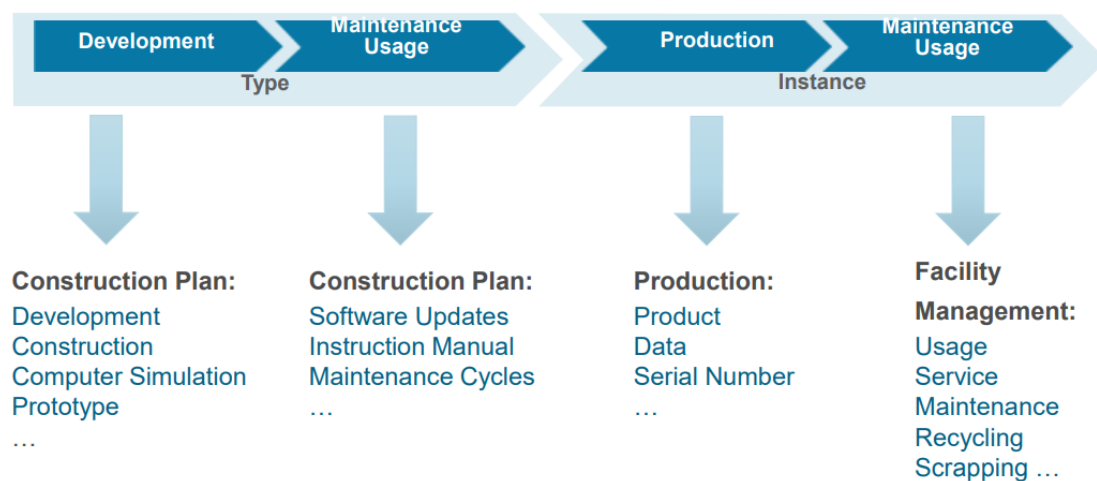


Figure 50. Product Life Cycle axis.

- Axis 3: Architecture

The six layers on the vertical axis describe the decomposition of a machine into its properties, structured layer by layer, i.e., the virtual mapping of a machine. Such representations originate from information and communication technology, where properties of complex systems are commonly broken down into layers (Figure 51).

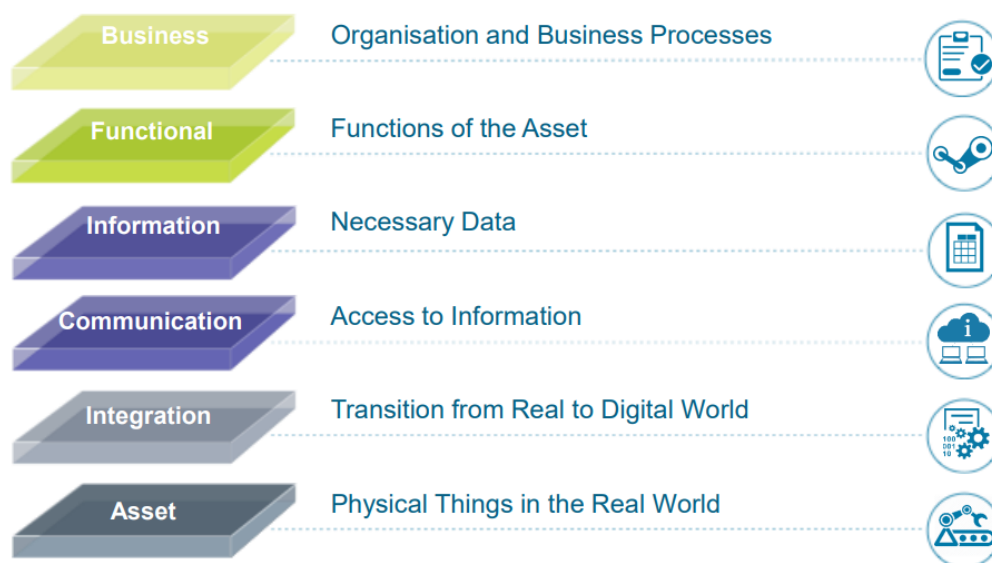


Figure 51. Architecture layers.

Within these three axes, all crucial aspects of Industry 4.0 can be mapped, allowing objects such as machines to be classified according to the model. Highly flexible Industry 4.0 concepts can thus be described and implemented using RAMI 4.0. The model allows for step-by-step migration from the present into the world of Industry 4.0.

Currently, others similar reference architecture models can be found for Industry 4.0, such as **Industrial Internet Reference Architecture (IIRA)** [IECC-4]. IIRA is a cross-industry reference architecture which

relates to a wide range of industries including energy, healthcare, manufacturing, and transportations. Similar to RAMI 4.0, It provides a five-layer description of the functions in an industrial system, their interrelation, structure and interactions. As it can be seen in Figure 52, both architecture models offer viewpoints that begin from the physical world and real-time data acquisition (Arrows 1,2,3) to higher level manufacturing controls (Arrow 4), data analytics (Arrows 5, 6), services and APIs (Arrows 7, 8), and business operations (Arrow 9).

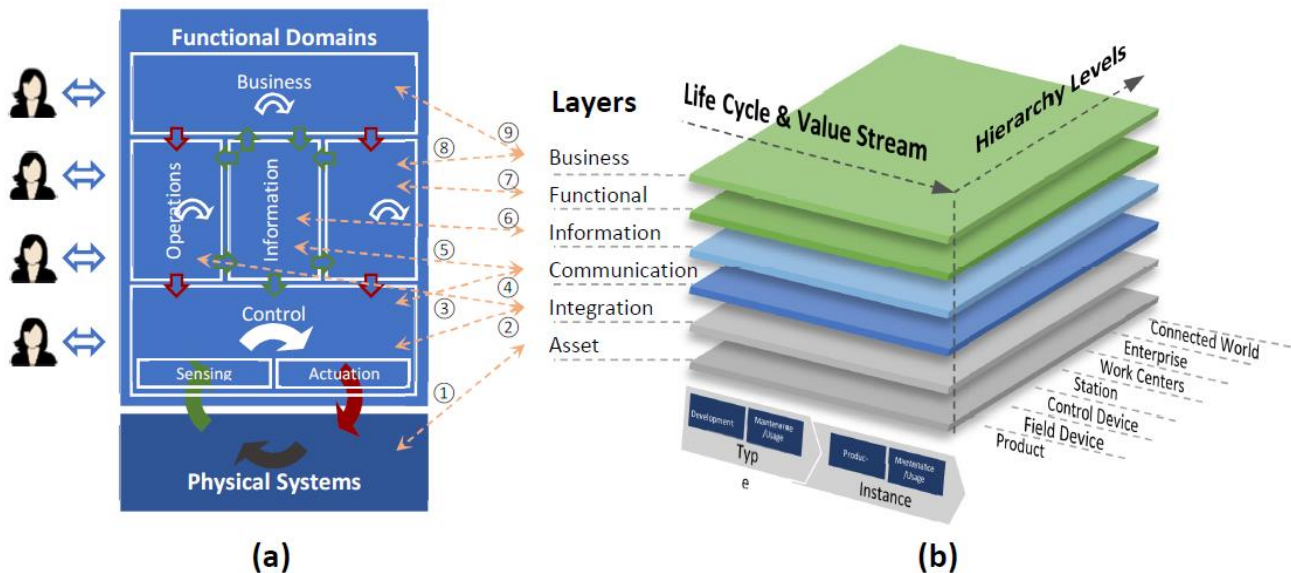


Figure 52. Functional viewpoint of IIRA (a) and RAMI4.0 (b)

On the other hand, IBM recently published a two-layer reference architecture for Industry 4.0, based on:

- *Platform/Hybrid cloud layer*: plant-wide data processing and analytics are performed, being sent, at the same time, to the enterprise layer. Commands are then sent back down to the edge, providing similar functions than previously discussed reference architectures, but with a broader scope and utilizing data from multiple plants.
- *Equipment/device layer*: Utilizing the edge, it is responsible from receiving data form physical devices, providing basic analytics and determining which information is sent to the higher levels. It sends commands to the smart devices at the same time.

3.3.1.2.2. Open Industry 4.0 Alliance for Industry 4.0

The aim of the Open Industry 4.0 Alliance (OI4.0) is to institute an alliance of innovative asset manufacturers (including asset digitization enablers) that adopts standards-based common semantic data models to enable the immediate instrumentation of smart assets in the end-to-end production life cycle of an operator, while bringing together the required critical mass of industry players. The vision of OI4.0 is to simplify the deployment and integration of intelligent assets into the operations of an operator (the end user, e.g., a factory) to a near “plug-and-play” level and provide pre-integrated high value solutions from Alliance members that can operate with operator-desired architecture openness.

An architecture that is presented by the Open Industry 4.0 Alliance and implemented by its members appears advantageous and is sketched prototypically: an open, scalable ecosystem with the following layers:

- Edge Connectivity (to the world of physical things).
- Edge Computing.
- Operator Cloud.
- A central repository for asset information and semantics.

Key principles are open interfaces, an open edge application layer and cloud application layer for the operator of a facility (either locally or in the cloud), data custodianship, role-based authorization for data access and private data and algorithms at every level for each provider and subscriber.

The Open Industry 4.0 alliance has designed a holistic architecture framework, keeping in mind to embrace all the important industry 4.0 standards and protocols:

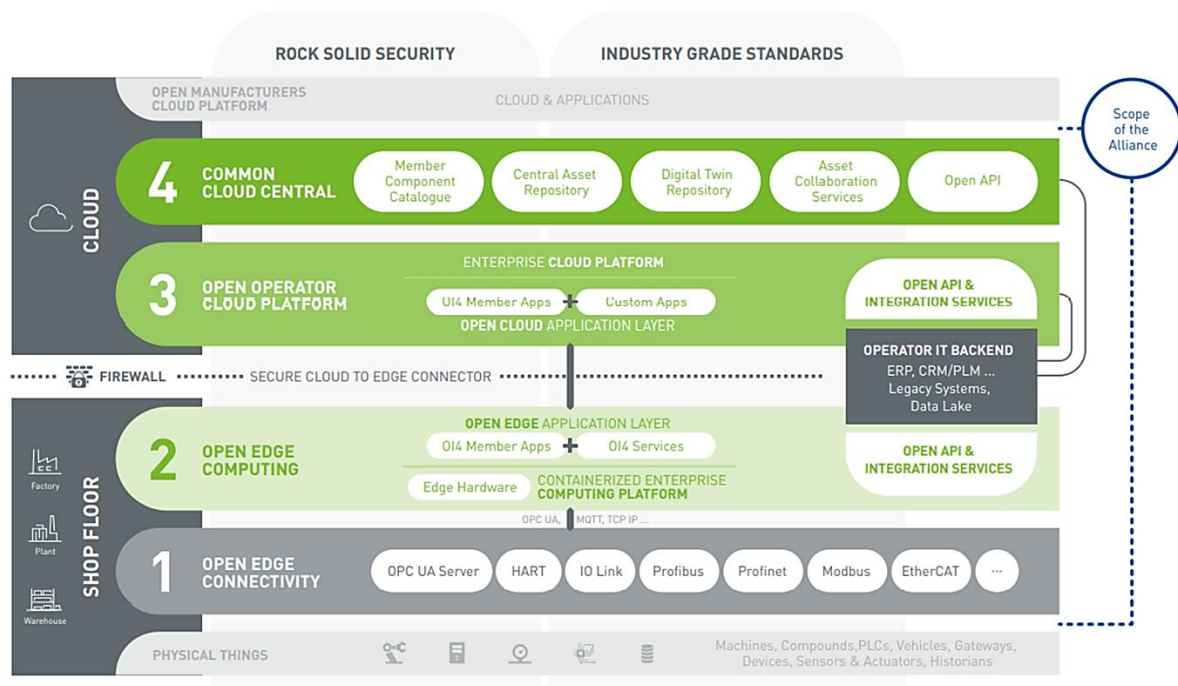


Figure 53. The open Industry 4.0 Alliance Technical Architecture [IECC-9].

As it can be seen in the Figure 53, the architecture comprises 4 main building blocks or layers:

1. **Open Edge Connectivity:** The open edge connectivity layer covers a wide range of possible data sources and possible communication technologies used.

Enable greenfield or brownfield connectivity scenarios for:

- Device (asset) identification
- Data conversion to compatible open edge computing platforms (e.g. MQTT, OPC UA, etc.)
- Local diagnostics

For connectivity of brownfield devices with analog protocols:

- Enable conversion of analog to digital protocols (e.g. using HART or IO Link)

2. **Open-Edge Computing:** Provides local data processing and an applications platform for plant operators, supervisors, warehouse users, etc. for real-time information about operational performance statistics. Edge computing is an emerging trend that provides direct access to applications for the users/operators of the machines.

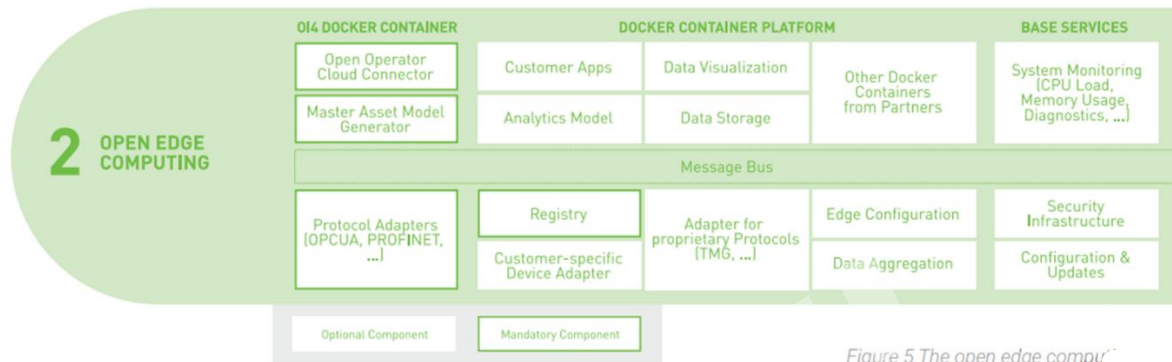


Figure 54 The open edge computing layer.

Figure 54. The open edge computing layer.

On the following lines, each of the mandatory components of the open edge computing layer will be discussed:

Open Operator Cloud Connector: In the OI4 architecture framework, every edge computing solution is expected to have a north-bound interface to communicate with the open operator cloud platform chosen by the operator. This component exposes an OI4-compliant communication and information model interface towards the message bus for tasks like device onboarding, while offering a platform-compatible interface to the Open Operator Cloud Platform layer.

Master Asset Model Generator: In order to allow identification and handling of assets in the OI4 architecture, each asset has to be assigned both an OI4.0 identifier and a master asset model. This component is responsible for generating these critical pieces of data for each asset being onboarded.

Protocol Adapters: In order to access the diverse and heterogeneous communication technologies on the open edge connectivity layer, a range of protocol adapters will have to be provided in the form of OI4.0 containers. These protocol adapters have the responsibility to encapsulate OT (operational technology) access both for onboarding and data acquisition tasks as well as any other access to the OT network they were written for that is requested over the message bus.

Registry: The registry has the critical task of keeping track of all onboarded assets as well as all containers deployed on the particular open edge computing platform. It serves as a directory of available entities to be addressed through appropriate topic structures in the message bus.

3. **Open Operator Cloud Platform:** Designed for enabling a trust-based environment, which would also provide consistent E2E interoperability and achieves the goal of faster adoption. The operator cloud as an IIoT platform should have all basic technical modules, e.g., device management and diagnostic, application enablement tools, data storage and processing, E2E security concepts, user management etc., as depicted in the solution.

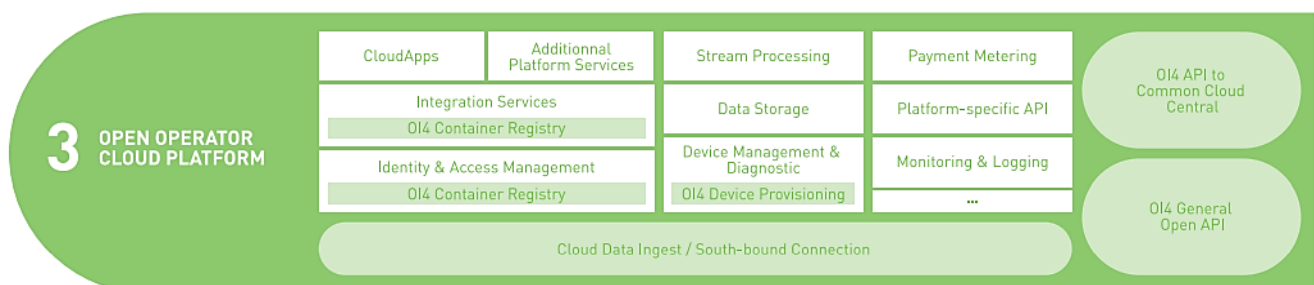


Figure 55. Open operator cloud.

The main technical modules of the Open Operator Cloud platform are described in the following lines:

OI4 API to Common Cloud Central: This API provides a standardized interaction between the Cloud Apps and Additional Platform Services of the Open Operator Cloud Platform and the Common Cloud Central of the Open Industry 4.0 Alliance.

OI4 General Open API: The module OI4 General Open API standardizes across Open Industry 4.0 Alliance members a secure access to other modules of the Open Operator Cloud, e.g., the access to the OI4 container registry or OI4 compliant device information.

Depending on the business and technology strategy of a company, there are several choices for the foundation of the Open Operator Cloud Platform. The Open Operator Cloud Platform may be based on one of the following scenarios or any composition of them:

- Operator-side or datacentre-located IIoT platforms as private or hybrid cloud, based on bare-metal or virtualized infrastructure
- Operator Cloud platform based on hyperscalers (e.g., Microsoft Azure, Amazon Web Services, Google Cloud Services, ...) or other highly scalable cloud infrastructure services
- Operator Cloud platform based on IIoT platform offerings provided by vendors or service providers for specific industrial use cases.

Cloud Data Ingest/South-bound connection: Data from multiple connected edge gateways can be ingested to the operator cloud via the technical module Cloud data ingest. Using this endpoint data is forwarded to further cloud components such as the module Stream processing or the module Data storage.

4. **Common Cloud Central:** use of a common cloud central layer as the main interoperability component by using a central asset information system to create a standardized semantic model. This enables the adoption of common data semantics in both the open operator cloud and open edge computing layers. It also helps to standardize and simplify application development efforts.



Figure 56. The common cloud central

The core elements of the common cloud central are structured as follows:

OI4 component catalogue: In order to allow acquisition and utilization of application functions in an OI4 context, containers have to be loaded into the Open Edge Computing and Open Operator Cloud Platform layers. In addition, operators have to have an overview of existing solutions in OI4 compliant devices and solutions.

Semantic Models Repository: The Semantic Models Repository will allow computerized access to the supported domain information standards. Thereby, even assets not fully covered by detailed type descriptions can be utilized and interpreted.

Type Information Repository: In order to allow the best effect of the Common Cloud Central platform of the Open Industry 4.0 Alliance, asset manufacturers are to supply information on the products they sell. This information is type specific and serves as a template for the Asset Administration Shells of concrete pieces of equipment. The information provided by manufacturers will be accessed through the Type Information Repository.

Instance Information Repository: In an Instance Information Repository, Digital Twins of assets are maintained that allow referencing asset instances and look up any historical information about the asset's lifecycle. This is a central component for the added value of OI4, as it makes asset information persistent beyond organizational boundaries. Due to this cross-boundary use case, access rights of the Instance Information Repository differ from those to the Type Information Repository. The semantics of the Instance Information Repository are modelled after those given in the Type Information Repository. Common cloud central will be part of a semantic network of asset information. For any one operator cloud, the common cloud central it addresses shall be unique. Differing providers of common cloud central services will have to interact in order to allow full information accessibility over all platforms. However, especially for brownfield use cases, where no information on a type might be present, the Semantic Models Repository can be utilized as a substitute for a baseline model. Hence, from the point of view of the Instance Information Repository, both the Type Information Repository and the Semantic Models Repository fulfil the same role.

Asset Network: The Asset Network structures the interactions of Asset Administration Shells that represent asset types and instances on the Common Cloud Central layer. The Asset Network in this allows the business processes of the Common Cloud Central layer and also other layers to fulfil their interaction needs with the repositories present on this layer. The data custodianship concept of the Open Industry 4.0 Alliance will be implemented through the Asset Network services.

Security concept of the OI4.0 Reference Architecture: From the security point of view, OI4 members are jointly working on various aspects of industrial cybersecurity in order to develop sustainable high value security concepts for customers use cases:

- Security-by-design: The pragmatic nature of the OI4 Alliance also underlies the OI4 cybersecurity workgroup. Security concepts are elaborated and tested closely together with the technical OI4 workgroups.
- The OI4 Alliance asserts a clear and comprehensible security concept. Every chosen technology in the OI4 ecosystem must meet the state-of-the-art requirements of security for encryption, authentication, data protection, and data privacy.
- The subject of industrial cybersecurity is considered holistically in the OI4 Alliance. Vertical and horizontal deep dives along the IIoT ecosystem are handled dynamically upon request or based on a specific use case relevance.

3.3.1.2.3. IoT-Edge data space continuum Approaches

One of the most important European initiatives regarding data spaces is The International Data Spaces Association (IDSA). IDSA is a coalition of more than 130 member companies that share a vision of a world where all companies self-determine usage rules and realize the full value of their data in secure, trusted, equal partnerships.

IDSA aims to reach a global standard for international data spaces (IDS) and interfaces, as well as fostering the related technologies and business models that will drive the data economy of the future across industries.

Recently, IDSA has designed the IDSA Reference Architecture Model (IDSA-RAM) in order to establish a reference framework for data spaces [IECC-8].

IDSA Reference Architecture Model aims at meeting the following requirements:

- **Trust:** Trust is the basis of the International Data Spaces. Each participant is evaluated and certified before being granted access to the trusted business ecosystem.
- **Security and data sovereignty:** Security is mainly ensured by the evaluation and certification of each technical component used in the International Data Spaces. In line with the central aspect of ensuring data sovereignty, a data owner in the International Data Spaces attaches usage restriction information to their data before it is transferred to a data consumer. To use the data, the data consumer must fully accept the data owner's usage policy.
- **Ecosystem of data:** It pursues the idea of decentralization of data storage, which means that data physically remains with the respective data owner until it is transferred to a trusted party. This approach requires a comprehensive description of each data source and the value and usability of data for other companies, combined with the ability to integrate domain-specific data vocabularies. In addition, brokers in the ecosystem provide services for real-time data search.
- **Standardized interoperability:** The International Data Spaces Connector, being a central component of the architecture, is implemented in different variants and can be acquired from different vendors. Nevertheless, each Connector is able to communicate with any other Connector (or other technical component) in the ecosystem of the International Data Space.
- **Value adding apps:** The International Data Spaces allows to inject apps into the IDS Connectors in order to provide services on top of data exchange processes. This includes services for data processing, data format alignment, and data exchange protocols, for example. Furthermore, data analytics services can be provided by remote execution of algorithms.

The reference architecture is shown in Figure 57, which is based in 5 general layers: Business, Functional, Process, Information and Systems. In addition, it comprises three perspectives that need to be implemented across all five layers: Security, Certification and Governance.

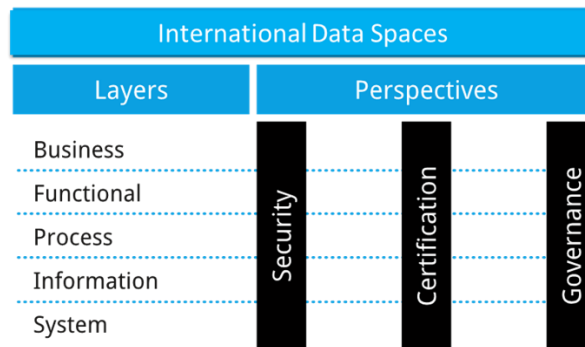


Figure 57. IDSA-RAM general structure [IECC-8].

Each of the **layers** of IDSA-RAM are briefly discussed in the following lines:

- **Business:** The Business Layer of the Reference Architecture Model defines and categorizes the different roles the participants in the International Data Spaces may assume. Furthermore, it specifies basic patterns of interaction taking place between these roles. It thereby contributes to the development of innovative business models and digital, data-driven services to be used by the participants in the International Data Spaces. While the Business Layer provides an abstract description of the roles in the International Data Spaces, it can be considered a blueprint for the other, more technical layers.
- **Functional:** It defines the functional requirements and the values to be implemented. The functional layer subdivides the requirements into six groups of software functionality to be provided by the IDS, which comply with the strategic goals shown previously:
 - *Trust:* The Trust group comprises three main aspects (roles, identity management, and user certification), which are complemented by governance aspects
 - *Security & Data Sovereignty:* The Security and data sovereignty group contains four major aspects: authentication authorization; usage policies usage enforcement; trustworthy communication security by design; and technical certification.
 - *Ecosystem of Data:* Being able to describe, find and correctly interpret data is another key aspect of the International Data Spaces. The Ecosystem of Data group comprises three major aspects: data source description, brokering, and vocabularies.
 - *Standardized interoperability:* Standardized data exchange between participants is the fundamental aspect of the International Data Spaces. The IDS Connector is the main technical component for this purpose.
 - *Value adding apps:* Before or after the actual data exchange, data may need to be processed or transformed. For this purpose, the International Data Spaces offers Data Apps. Each Data App has a lifecycle, spanning its implementation, provision in the App Store, installation, and support.
 - *Data markets:* Data to be exchanged in the International Data Spaces may have monetary value. Therefore, the International Data Spaces has to integrate data market concepts, like clearing and billing, but also governance.
- **Process:** The Process Layer specifies the interactions taking place between the different components of the International Data Spaces. It thereby provides a dynamic view of the Reference Architecture Model:
 - *Onboarding*, i.e., what to do to be granted access to the International Data Spaces as a Data Provider or Data Consumer.
 - *Data Offering*, i.e., offering data or searching for a suitable data.
 - *Contract Negotiation*, i.e., accept data offers by negotiating the usage policies.
 - *Exchanging Data*, i.e., transfer data between IDS Participants.

- *Publishing and using Data Apps*, i.e., interacting with an IDS App Store or using IDS Data Apps.
- **Information:** The Information Layer specifies the Information Model, the domain-agnostic, common language of the International Data Spaces. The Information Model is an essential agreement shared by the participants and components of the IDS, facilitating compatibility and interoperability. The primary purpose of this formal model is to enable (semi-)automated exchange of digital resources within a trusted ecosystem of distributed parties, while preserving data sovereignty of Data Owners. The Information Model therefore supports the description, publication and identification of data products and reusable data processing software (both referred to hereinafter as Digital Resources, or simply Resources). Once the relevant Resources are identified, they can be exchanged and consumed via easily discoverable services. Apart from those core commodities, the Information Model describes essential constituents of the International Data Spaces, its participants, its infrastructure components, and its processes.

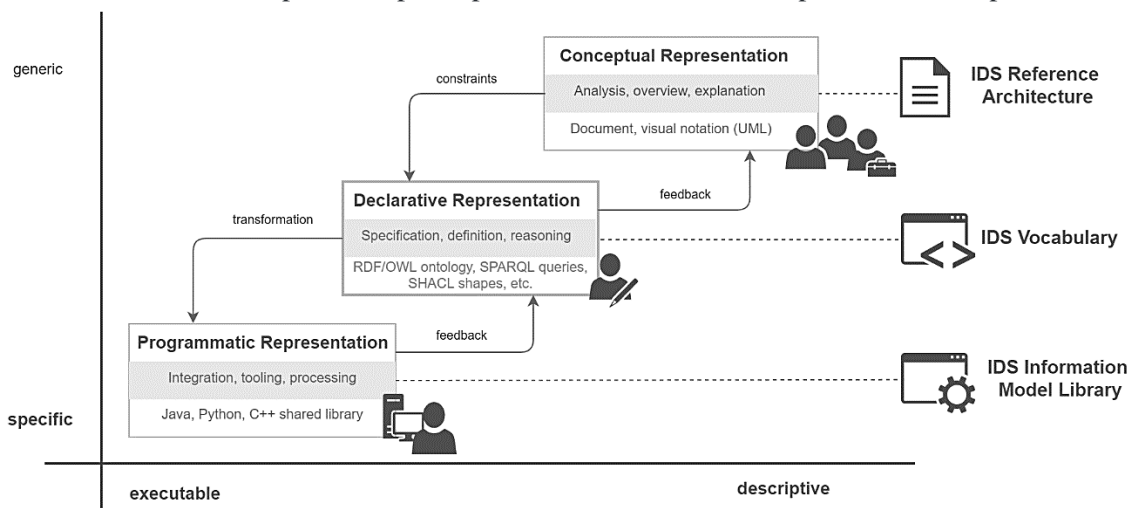


Figure 58. Representation of the information model of IDSA-RAM [IECC-8].

The three cross-sectional **perspectives** are directly related to the five layers of the IDSA-RAM, which are further detailed in the following lines:

- **Security:** As discussed previously, one strategic requirement of the IDS is to provide secure data supply chains. The IDS Security Architecture provides means to identify devices in the IDS, protect communication and data exchange transactions, and control the use of data after it has been exchanged. To control the use of data, Access Control restricts access to resources. Authorization is the process of granting permission to resources. There are several models of Access Control, such as Discretionary Access Control (DAC), Mandatory Access Control (MAC), Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), etc. RBAC and ABAC are the most frequently used models.
- **Certification:** As stated previously, data sovereignty is one of the fundamental value propositions for the IDS. Data sovereignty can be defined as a natural person's or legal entity's capability of being in full control of its data. To enable this control, each participant needs to follow the agreed rules for the IDS and requires reliable information about the guarantees offered by potential business partners. This requires a certification based on different roles: applicants, evaluation facilities and certification body. It consists of collecting evidence in form of documentation and interviews with employees in four different assessments: Quality Management System, Security Management System, Competence of the Evaluators and Testing equipment and its usage (only relevant for Component Certification).
- **Governance:** It defines the roles, functions, and processes of the International Data Spaces from a governance and compliance point of view. It thereby defines the requirements to be met by the business ecosystem to achieve secure and reliable corporate interoperability. The International Data Spaces supports governance issues by: (i) Providing an infrastructure for data exchange, corporate interoperability, and the use of new, digital business models, (ii) Establishing trustworthy relationships between Data Owners, Data Providers, and Data Consumers, (iii) Acting as a trustee for mediation between participants, (iv) Facilitating negotiation of agreements and contracts, (v) Aiming at transparency and traceability of data exchange and data use, (vi) Allowing private and public data

exchange, (vii) Taking into account individual requirements of the participants and (viii) Offering a decentralized architecture that does not require a central authority.

By proposing an architecture for secure data exchange and trusted data sharing, the International Data Spaces contributes to the design of enterprise architectures in commercial and industrial digitization scenarios. It does so by bridging the gaps between research, industrial stakeholders, political stakeholders, and standards bodies.

3.3.1.2.4. Fiware Smart Industry Reference Architecture

Together with its members and partners, FIWARE Foundation drives the definition – and the Open Source implementation – of key open standards that enable the development of portable and interoperable smart solutions in a faster, easier and affordable way, avoiding vendor lock-in scenarios, whilst also nurturing FIWARE as a sustainable and innovation-driven business ecosystem. Keeping freedom in decision making, openness, transparency and meritocracy are the cornerstones and principles of the FIWARE Community. An important factor of the “FIWARE Culture” driving innovation and performance is the balanced collaboration between individuals who invest time and effort, companies that build businesses with and on FIWARE, and the researchers, developers and integrators who develop and deploy new applications based on FIWARE technologies. FIWARE’s open-source development and business empower communities from different sectors, backgrounds and geographies to contribute and co-create. It’s the FIWARE ecosystem that makes the FIWARE users successful. Among FIWARE’s ecosystem, the FIWARE’S Smart Industry Reference Architecture (Figure 59) enables smart industry applications:

- Building a smart manufacturing platform, based on standards and other open source components, that support real-time, high-value applications to optimize production systems and value chains.
- Creating a reference architecture, compliant with existing industry architectures such as the Reference Architecture Model Industrie 4.0, the Industrial Data Space Reference Architecture or the Industrial Internet Consortium Reference Architecture which are capable of transforming the industrial sector into a networked, data-driven environment.
- Breaking the information silos and unleashing the potential of context data from the Internet of Things and different systems, which can be exploited together using Big Data and Artificial Intelligence services on the Cloud to achieve higher degrees of efficiency and automation.
- Using a data-driven approach through the decoupling of industrial processes while warranting sovereignty on a strategic asset: data.

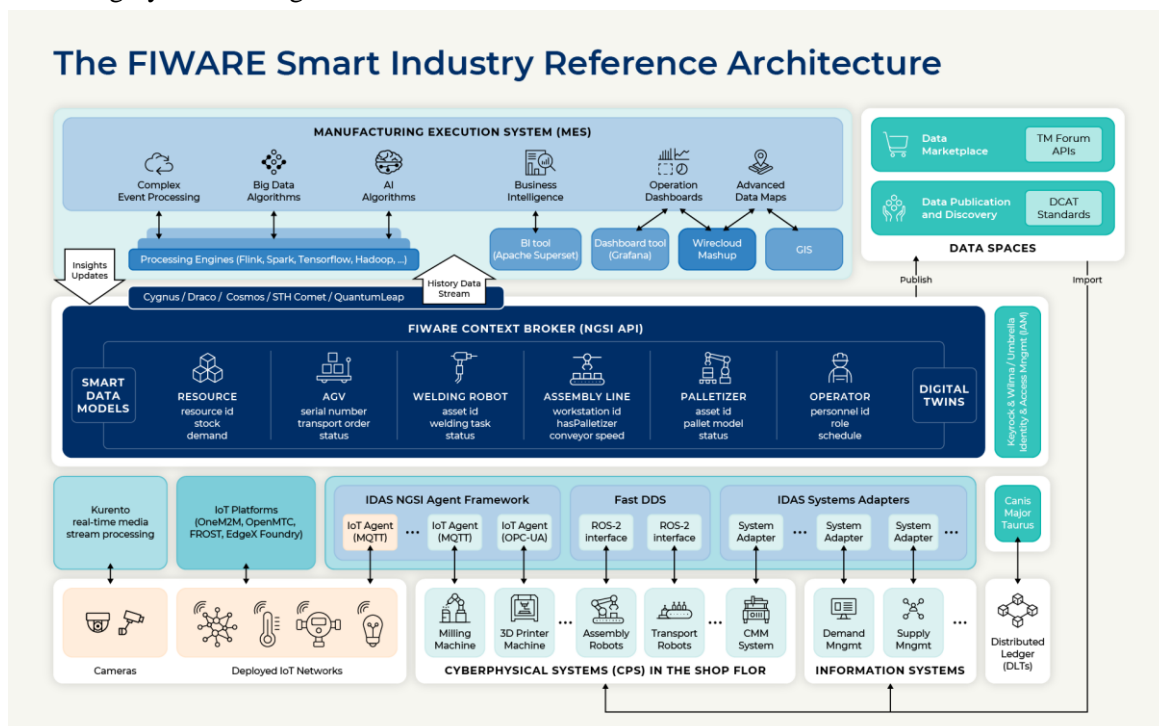


Figure 59. FIWARE Smart Industry Reference Architecture [IECC-4]

FIWARE reference architecture technology can be used in key scenarios [IECC-4]:

- A Machine Data Bus is the real-world data exchange, active on the factory shop floor, in the operations of a product or in transportation and logistics
- A Unit Data Bus uses dedicated edge/fog data gateways as a bridge between real and digital worlds. Here, a cloud-edge programming system can process the data stream through a series of distributed components, using the **FIWARE NGSI API** to harmonize access to data published using many different data formats.
- A Site Data Bus implements the data exchange in a single administrative domain, be it a company, an IT department, a plant or a fleet of vehicles. This would make use of **FIWARE Context Broker** technology for managing the entire lifecycle of context information including updates, queries, registrations, and subscriptions.
- An Inter-site Data Bus materializes B2B data exchange and sharing of data between business processes distributed across at least two different administrative domains. FIWARE Context Broker technology can also be used for this purpose.

Regarding the use of data space, previously presented IDSA and FIARE have a shared vision, as they are currently working together on the first open-source implementation of the IDS Reference Architecture. Its main component is the IDS Connector which, based on the FIWARE Context Broker and other complementary FIWARE technologies, manages all aspects related to the publication of and the access to data. Both the IDS and FIWARE platforms are listed as promising digital industrial platforms build on European strength in a recent report published by the European Union on the progress of the Digitising European Industry (DEI) initiative [10].

The core communication component of an IDS Connector implemented using FIWARE is the FIWARE Context Broker component (Orion). Orion Context Broker comes together with components enabling:

- Enforcement of data usage control policies: Wilma (PEP)
- Federation with Context Brokers associated to remote IDS Connectors
- Accounting of interactions (requests, notifications): Wilma (CDR gen)
- Connection with alternative processing engines or data sinks: Cygnus

Tools enable the automated deployment of data system adapters or data processing engines and configure connections to preserve defined policies. Authorization and Access Control components adhere to widely accepted open standards (XACML: PEP + PDP/PAP) while automated deployment tools rely on latest developments with Docker or Kubernetes.

3.3.1.2.5. Data-driven DFA reference models for zero-x manufacturing

The digital factory alliance (DFA) is born under the umbrella of European Commission projects aiming at modernizing and digitalizing the assets of the factories of the future, with the strong conviction that these actions will have a critical influence in the way these factories will be operated and managed in the years to come, by promoting the use of Artificial Intelligence Technologies and Data Intelligence to strive for Zero X Manufacturing Environments.

This initiative allows its members to get access to the most updated knowledge, trends and “ready-to-deploy” products in the digital manufacturing field, gaining exposure to a growing Zero X Manufacturing marketplace, with the added brand recognition and access to new business opportunities. The DFA also provides an opportunity to participate in unique business networks that will allow its participants to quickly and effectively respond to crisis scenarios and critical manufacturing demands where supply chains are compromised, gaining resilience and the capacity to keep operating in repurposed manufacturing scenarios.

DFA Reference Models is the starting point towards digital service integration. The DFA provides a common framework for integration of digital products and data-driven service platforms. The DFA provides a unified approach to gradual digital transformation based on adoption of secure Industrial IoT, Big Data Analytics, Artificial Intelligence & Machine Learning, Edge Computing and Digital Twin technologies.

As discussed in the previous section, the industry 4.0 initiative proposed the digital transformation of European factories towards smart digital production systems through intense vertical and horizontal integration, with the objective to increase operational efficiency, scrap reduction, prescriptive quality management, energy efficiency, defect avoidance and improved smart product customer experience, fostering new digital business models. This demands for the definition of reference models and system architectural approaches that could help to manage the complexity of this revolution. The challenges developing the Reference Architecture for digital Zero-Defect Manufacturing (ZDM) solutions for smart manufacturing, based on relevant sector standards and adopting the most mature innovative technologies for digital manufacturing, based on innovative technologies and on relevant sector standards such as RAMI 4.0, as it is shown in Figure 60.

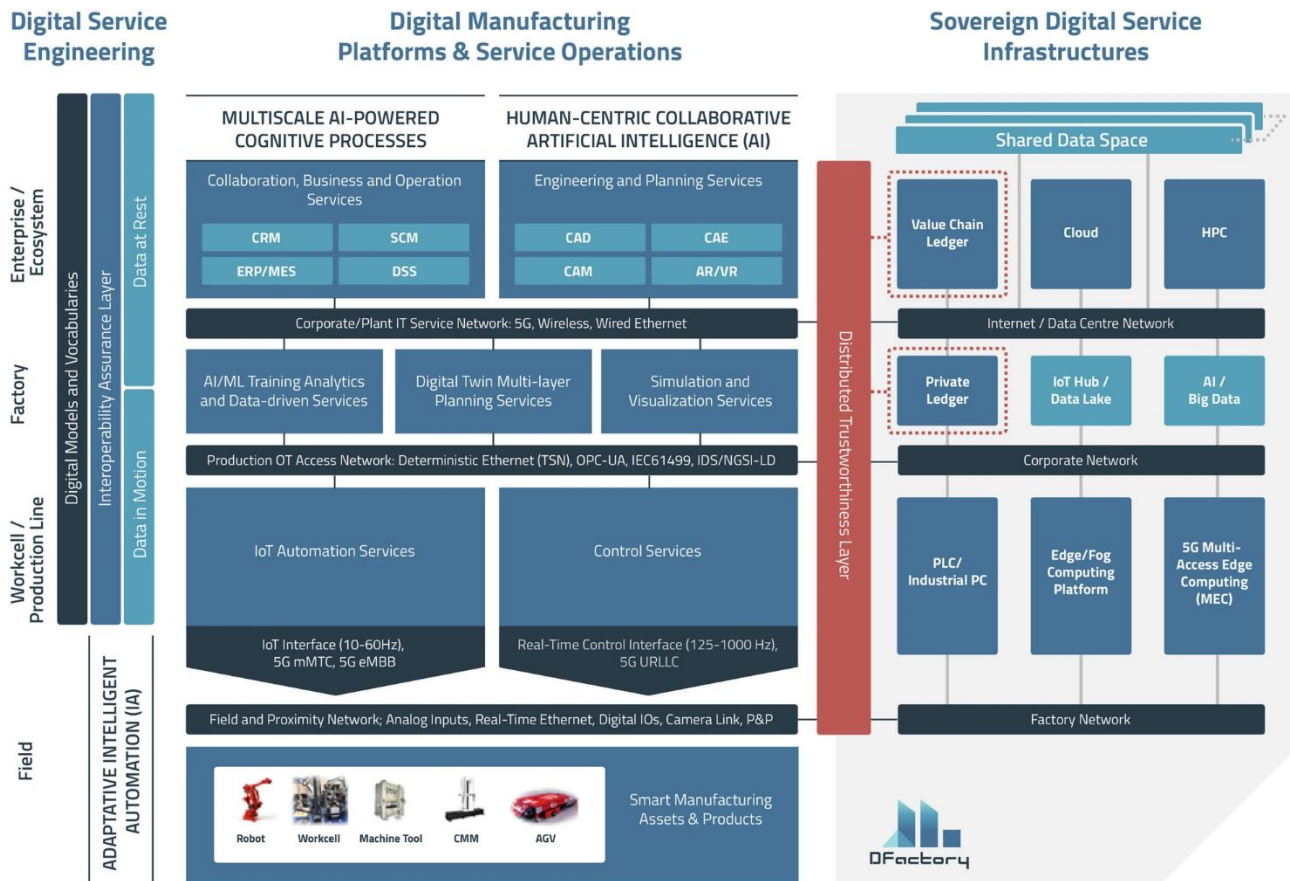


Figure 60. Digital Factory Alliance Reference Architecture for Industry 4.0 [IECC-7]

The DFA SD-RA design complies with ISO/IEC/IEEE 42010 [12] architectural design principles and provides an integrated yet manageable view of digital factory services. In fact, DFA SD-RA integrates functional, information, networking and system deployment views under one unified framework. The DFA SD-RA address the need for an integrated approach to how (autonomous) services can be engineered, deployed and operated/optimized in the context of the digital factory. With this aim, the DFA SD-RA is composed of three main pillars, as depicted in Figure 60:

- **Digital Service Engineering.** This pillar provides the capability in the architecture to support collaborative model-based service enterprise approaches to digital service engineering of (autonomous) data-driven processes with a focus on supporting smart digital engineering and smart digital planning and commissioning solutions to the digital factory. The pillar is mainly concerned with the harmonization of digital models and vocabularies. It is this pillar that should develop interoperability assurance layer capabilities with a focus on mature digital factory standards adoption and evolution towards an “industry commons” approach for acceleration of big data integration, processing and management. It is this pillar where “security by design” can be applied both at the big data, manufacturing process and shared data space levels.

- **Digital Manufacturing Platforms and Service Operations.** This pillar supports the deployment of services and DMPs across the different layers of the digital factory to enact data-driven smart digital workplaces, smart connected production and smart service and maintenance manufacturing processes. The pillar is fundamental in the development of three enabling capabilities central to the gradual evolution of autonomy in advanced manufacturing processes, i.e. multi-scale AI-powered cognitive processes, human-centric collaborative intelligence and adaptive Intelligent Automation (IA). The enablement of both knowledge-based (multi-scale artificial intelligence) and data-driven approaches (collaborative intelligence) to digital factory intelligence is facilitated by the support of service-oriented and event-driven architectures (interconnected OT and IT interworking event and data buses) embracing international and common standard data models and open APIs, thereby enabling enhanced automated context development and management for advanced data-driven decision support.
- **Sovereign Digital Service Infrastructures.** The operation of advanced digital engineering and digital manufacturing platforms relies on the availability of suitable digital infrastructures and the ability to effectively develop a digital thread within and across the digital factory value chain. DFA SD-RA relies on infrastructure federation and sovereignty as the main design principles for the development of the data-driven architecture. This pillar is responsible for capturing the different digital computing infrastructures that need to be resiliently networked and orchestrated to support the development of different levels and types of intelligence across the digital factory. In particular, the DFA SD-RA considers three main networking domains for big data service operation; i.e. factory, corporate and internet domain. Each of these domains needs to be equipped with a suitable security and safety level so that a seamless and cross-domain distributed and trustworthy computing continuum can be realized. The pillar considers from factory-level digital infrastructure deployment such as PLC, industrial PC or Fog/Edge to the deployment of telecom-managed infrastructure such as 5G multi-access edge computing platforms (MEP). At the corporate level, the reference architecture addresses the need for the development of IoT Hubs that are able to process continuous data streams as well as dedicated big data lake infrastructures, where batch processing and advanced analytic/learning services can be implemented. It is at this corporate level that private ledger infrastructures are unveiled. Finally, at the internet or data centre level, the digital factory deploys advanced computing infrastructures exploiting HPC, Cloud or value chain ledger infrastructures that interact with the federated and shared data spaces.

The DFA RA is aligned with ISO 20547 Big Data Reference Architecture. The DFA Sovereign Digital Service Infrastructures pillar allows reference model to additionally address the ISO 20547 Big Data Framework Provider layer. The DFA RA is composed of four layers [11] that address the implementation of the 6 big data “C” (Connection, Cloud/edge, Cyber, Context, Community, Customization), enables four different types of intelligence (smart asset functioning, reactive reasoning, deliberative reasoning and collaborative decision support) to be orchestrated and maps to the 6 layers of the RAMI 4.0 (product, devices, station, WorkCentre, enterprise and connected world), which target all relevant layers required for the implementation of AI-powered data-driven digital manufacturing processes:

- 1- The lower layer of the DFA RA contains the field devices in the shopfloor: machines, robots, conveyer belts as well as controllers, sensors and actuators are positioned. Also in this layer the smart product would be placed. This layer is responsible for supporting the development of different levels of autonomy and smart product and device (asset) services leveraging on intelligent automation and self-adaptive manufacturing asset capabilities.
- 2- The workcell/production line layer represents the individual production line or cell within a factory, which includes individual machines, robots, etc. It covers both the services, that can be grouped in two those that provide information about the process and the conditions (IoT automation services), and the actuation and control services (automation control services); and the infrastructure, typically represented in the form of PLC, industrial PCs, edge and fog computing systems or managed telecom infrastructures such as MEC. This layer is responsible for developing reactive (fast) reasoning capabilities (automated decision) in the SD-RA and leveraging augmented distributed intelligence capacities based on enhanced management of context and cyber-physical production collaboration.
- 3- At the factory level, a single factory is depicted, including all the work cells or production lines available for the complete production, as well as the factory-specific infrastructure. Three kinds of services are typically mapped in this layer: (1) AI/ML training, analytics and data-driven services; (2) digital twin

multi-layer planning services; and (3) simulation and visualization services. The infrastructure that corresponds to this layer is the IoT Hubs, data lakes and AI and big data infrastructure. This layer is responsible for supporting the implementation of deliberative reasoning approaches in the digital factory with planning (analytical, predictive and prescriptive capabilities) and orchestration capabilities, which combine and optimize the use of analytical models (knowledge and physics based), machine learning (data-driven), high-fidelity simulation (complex physical model) and hybrid analytics (combining data-driven and model-based methods) under a unified computing framework. This leverages in the architecture collaborative assisted intelligence for explainable AI-driven decision processes in the manufacturing environment.

- 4- The higher layer refers to the enterprise/ecosystem level, that encompasses all enterprise and ecosystem (connected world) services, platforms and infrastructures as well as interaction with third parties (value chains) and other factories. The global software systems that are common to all the factories (collaboration business and operation services as well as engineering and planning services) are supported usually by Cloud or HPC infrastructures. It is this layer that supports the implementation of shared data spaces and value-chain-level distributed ledger infrastructures for implementation of trusted information exchange and federated processing across shared digital twins and asset administration shells (AAS). This layer leverages a human-centric augmented visualization and interaction capability in the context of data-driven advanced decision support or generative manufacturing process engineering.

As discussed in this section, the development of Industry 4.0 and Industry 5.0 concepts requires the design of architecture models that support technological enablers, such as low latency communication protocols, smart devices, edge computing and advanced automation, among others. Currently existing reference architectures, such as RAMI4.0, OI4.0 Reference Architecture and DFA reference model for zero x manufacturing, aim to face new industrial challenges, intending to act as I4.0 and I5.0 enablers.

3.3.2. Current existing standards related to aerOS

Developing a meta-operating system for the edge-cloud continuum, intended for several verticals such as industry or logistics, undoubtedly means adhering to well-defined international standards for technologies, protocols, and good practices. This section discusses some of the standards that, at the moment of publication of this report, are of relevance for the development and implementation of aerOS.

The standards and their developing organizations (SDOs) are classified by their technology areas as follows:

- Standards on Data Exchange and Modelling
- Standards on Networking and Communication Technologies

3.3.2.1. Standards on Data Exchange and Modelling

3.3.2.1.1. Data Distribution Service (DDS)

SDO: Object Management Group (OMG).

The Data Distribution Service (DDS) standard creates a framework for real-time data exchange between machines. It describes a Data-Centric Publish-Subscribe (DCPS) model that provides applications with a single interface to information generated and stored in a distributed manner. The standard specification summarizes itself as an enabler of the “Efficient and Robust Delivery of the Right Information to the Right Place at the Right Time.” [CES-1]

The goal of this technology is then the efficient delivery of information from producers and data-storage agents to matching consumers. To achieve that, the standard sets the guidelines for the creation of a communications middleware that handles all transfer functions. Once the middleware is executed by a device, it becomes a “node” in the DDS network. If it produces or stores some specific set of data, the device becomes a “publisher”, categorizes information in the form of “topics” and publishes it as “samples.” The devices that consume that data, in the other hand, are called “subscribers”, as they advertise to the network their intention to subscribe to topics. After that intention is acknowledged by the publisher in charge of the topic, this will send the samples to the subscriber at the rate described by the subscription request [CES-1].

Traditionally, the Pub/Sub messaging pattern relies on a Pub/Sub broker sitting in the network whose function is to match publishers and subscribers, however DDS removes that broker and replaces it with a virtual databus (inspired by databuses used in real-time fieldbus communication protocols). This allows for a better scalation of the network and the removal of a single point of failure. Both constraints of other Pub/Sub communication technologies. [CES-2]

3.3.2.1.2. OPC Unified Architecture (OPC UA)

SDO: OPC Foundation (Development), IEC (Publication).

Published as: IEC 62514.

OPC UA is a set of open-source standards for data exchange and modelling developed mainly for industrial use cases. It describes a platform-independent and service-oriented framework that aims to provide common data models to be used simultaneously by sensors and actuators in the factory floor (OT), as well as control, management, planning and accounting systems (IT). That reduces the logical constraints imposed by the division between those two technology realms, and allows for the use of one shared IP-based infrastructure, in contrast to several co-existent heterogeneous network technologies [CES-3].

Given its goal to put together all the traffic into a single IP-based infrastructure, OPC UA has become a very important complement of network technology standards that share a similar approach, such as TSN. However, OPC UA is independent from the underlying communication protocol. The standard provides mappings to several protocols, such as TCP/IP, UDP/IP, WebSockets, AMQP and MQTT [CES3].

Even though, the standard gives a great deal of freedom from the communication technology perspective, it addresses other aspects of the data exchange with detail, such as security, extensibility, platform independence, and access to information models [CES-4].

Two messaging patterns are considered by the OPC UA standards, client-server communication and Pub/Sub. The former is achieved via services provided by a server to the clients, following the design paradigm of service-oriented architecture. The latter, Pub/Sub, relies on a message-oriented middleware that acts as broker, handling the message exchange and decoupling publishers and subscribers [CES-4].

3.3.2.1.3. Yet Another Next Generation (YANG)

SDO: Internet Engineering Task Force (IETF).

Published as: IETF RFC 7950

The YANG standards describe a data modelling language developed mainly for network management protocols, such as NETCONF. Given the rapid industry adoption of NETCONF, YANG is very common in highly automated networks. It has become the de-facto standard language to describe attributes of network elements.

YANG provides a set of built-in data types as well as the capability to define custom types, thanks to its C-like syntax and hierarchical data organization. It emphasizes readability, modularity and flexibility. [CES-5]

A YANG module defines a single data model and determines its encoding. A module can be a complete, stand-alone entity, or it can reference definitions in other modules, as well as augment other data models with additional nodes. This allows for the creation of syntactic configuration data that meets constraint requirements and the validation of the data in the model before it is loaded and committed to a network device. [CES-5]

3.3.2.1.4. Next Generation Service Interface-Linked Data (NGSI-LD)

SDO: ETSI Industry Specification Group (ISG) Cross-cutting Context Information Management (CIM)

The ETSI CIM defines a standard for exchanging contextualized data between smart applications. To this end, the standard introduces the NGSI-LD protocol, which is composed of two innovations: the NGSI-LD information model and the NGSI-LD API.

The NGSI-LD information mode builds on the Labelled Property Graph (LPG) model, which has become the main option by popular graph databases like Neo4j. The LPG model defines entities that have relationships with other entities. In turn, both entities and relationships can have properties that provide additional characteristics.

The foundational classes of the LPG model are formally described in the NGSI-LD meta-model with an OWL ontology as depicted in Figure 1.

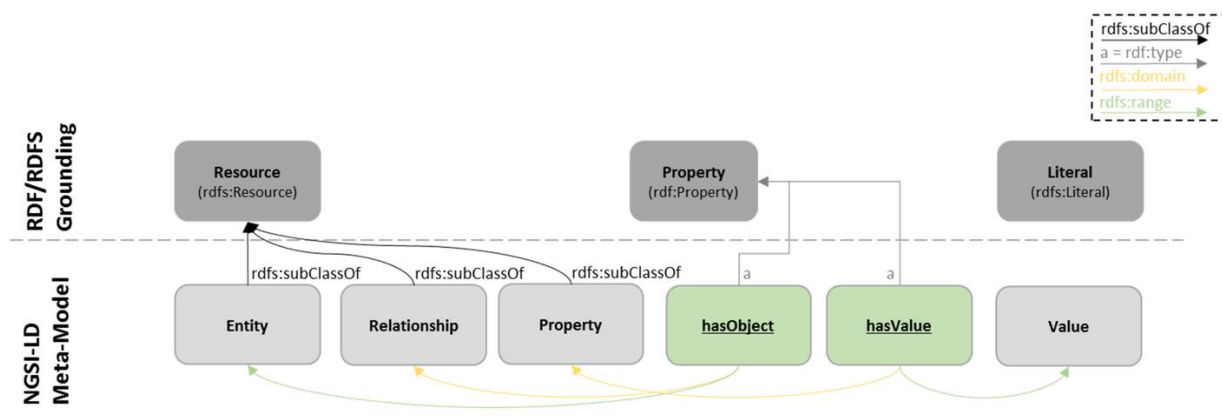


Figure 61. NGSI-LD meta-model [CES-6]

Grounding on the Semantic Web standards (e.g., RDF, RDFS, OWL) enables the NGSI-LD information model to map the NGSI-LD meta-model with higher level ontologies such as cross-domain or domain-specific ontologies. Therefore, NGSI-LD information models combine the best of both graph modelling approaches: (i) the compact, natural representation of property graphs; (ii) referencing public ontologies that can be leveraged for semantic reasoning. Additionally, the NGSI-LD information model extends the expressiveness of the LPG model by supporting the definition of properties-of-properties, relationship-of-relationships, and relationships-of-properties.

The second main innovation introduced by the standard is the NGSI-LD API [CES-7]. The NGSI-LD API implements a RESTful-based API for exchanging context information that follows the structure of the NGSI-LD information model. The standard does not define any specific architecture for this API, though some prototypical architectures are considered such as the distributed architecture depicted in Figure 2.

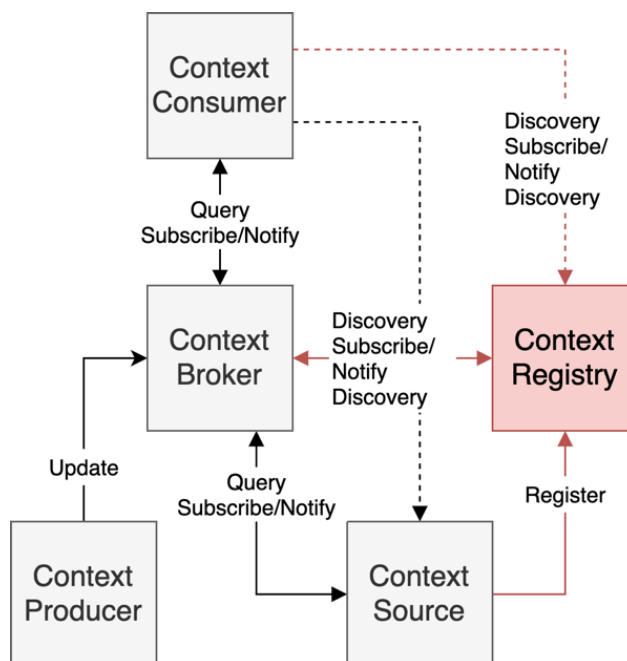


Figure 62. NGSI-LD distributed architecture

This architecture illustrates the different components and their interactions to exchange context information. The Context Broker represents the core of the architecture element and is responsible for storing context information and exposing it through the NGSI-LD API. At the bottom, we find two types of Context Providers: Context Producers and Context Sources. The role of the Context Producer is to write in the Context Broker context information that has been collected from a given data source. On the other hand, the role of the Context

Source is slightly more complex. This component registers in the Context Registry what kind of context information it can provide, so when the Context Broker needs such information, it forwards the request to the Context Source. Lastly, at the top of the architecture we find the Context Consumer. This role interacts with the Context Broker to search for context information or to subscribe for updates on context information, which can be sent either periodically or based on changes. Additionally, Context Consumers can inspect the contents of the Context Registry to find Context Sources, so then they can consume context information directly from the Context Source without having to go through the Context Broker.

3.3.2.1.5. Smart Data Models

Smart Data Models is a collaborative program aiming to allow data interchange between organizations providing multi-sector, agile-standardized, free and open-licensed data models based on current use cases and open standards.

Currently FIWARE Foundation, IUDX, TMForum, and OASC are its steering board members but more than other 70 organizations have already contributed to the program.

A data model is the description of any physical or logical entity used in any system. It lists their elements (attributes, fields, whatever you called them), the type of data each element stores and a text description of the attribute. This data models are coded using JSON Schema, a subset of JSON which is a declarative language that allows you to annotate and validate JSON documents.

Smart Data Models program provides data models for digital twins and data spaces on the following basis:

- Free and open-licensed data models for digital market (0€ cost)
- Multi-sector
- Based on real use cases and adopted open standards. Collaborative development.
- At market speed
- Customizable to local needs
- Compatible with linked data

The reason for this initiative comes from the need of systems on any organization to interchange data with external entities for many different purposes or just to make business with suppliers and clients. NGSI and NGSI-LD standards allows requesting data from many systems with a standard format, REST compatible, that can cope with most of the needs including geoquerying, next to real-time requests and heterogeneous sources. So, sharing data models can help to gain full interoperability between different systems providing the how to request for data and the structure of the retrieved data.

Currently, there are several Domains available in Smart Data Model, each of the domains contains their own data models. Each domain has its own github repository. There are 2 special repositories “incubated” and “harmonization”. The incubated repository is where new data models are contributed and tested before they can officially belong to one of the official domains. Another special repository is Harmonization, where data models are to be completed or to be harmonized with other data models or regulations.

Smart Cities	Smart Agrifood	Smart Water	Smart Energy	Smart Logistics
Smart Robotics	Smart Sensoring	Cross sector	Smart Health	Smart Destination
Smart Environment	Smart Aeronautics	Smart Manufacturing	Incubated	Harmonization

Figure 63. Repositories and domains in Smart data models

As explained previously, data models are contributed to the program by individuals or organizations on some well defined basis which imply the contribution of the data model with an Open licence. The data model must be based on the implementation of a real case scenario and must meet the defined code guide lines, the contribution must be consistent with the current naming of attributes. The contribution must provide the corresponding JSON-Schema and must provide an example on JSON or JSON-LD.

The Smart Data Models activity is based on the seven principles of agile standardization as a complementary approach to the classical standardization. It allows a very quick (days) definition of the data models, the documentation in 7 languages, searchable in a specific tool, and the generation of additional examples.

3.3.2.2. Standards on Networking and Communication Technologies

3.3.2.2.1. Time Sensitive Networking (TSN)

SDO: Institute of Electrical and Electronic Engineers (IEEE)

Published as: IEEE 802-

Time-sensitive networking (TSN) is a set of open standards that provide deterministic, reliable, high-bandwidth, low-latency communication [CES-8]. TSN is specified by IEEE 802 and aims to enable Ethernet networks to give QoS guarantees for time-sensitive and/or mission-critical traffic and applications. Different QoS assurances are offered by the different TSN standards. Profiles are being defined as devices from several suppliers must offer functional compatibility. To reduce the complexity that can be brought on by potential variations in standards, these profiles concentrate on a common set of functions and settings.

The functions standardized in the profiles, can be categorized in three main elements that constitute the complete TSN solution. Those are: Time synchronization, scheduling, and traffic shaping / path control.

Time synchronization is necessary to achieve determinism on a TSN network. Additionally, it allows the network to carry TSN scheduled traffic. The standard protocol for time synchronization in TSN is the IEEE 802.1AS generalized Precision Time Protocol (gPTP) [CES-9], which derives from the IEEE 1588 Precision Time Protocol (PTP) [CES-10] and allows for time synchronization over Ethernet only.

The gPTP synchronization process is described in IEEE 1588 [CES-10]: A central PTP instance (also called a “grand-master”), sends its current time information to all connected gPTP instances simultaneously, with the use of Ethernet multicast, for instance. With this information, the receiving gPTP instances adjust their clocks, correcting for the propagation time between them and the grand master. This propagation delay is continuously updated by measuring round trip times between the grand-master and each other gPTP instance.

Scheduled traffic is a time-based resource allocation mechanism, where traffic classes with different priorities are given different time windows to transmit on certain links, populating buffers mainly with lower-priority traffic. Those priorities are determined by looking at the priority code point (PCP) indicator in the VLAN tag of the Ethernet headers. This mechanism is standardized in IEEE 802.1Q [CES-11].

TSN also defines several traffic shaping and path control mechanisms. One of those is Frame pre-emption, where packets of higher priority pre-empt those of lower, to guarantee that the former traverse the network without much interference. This mechanism can be combined with traffic scheduling. Frame pre-emption is standardized in IEEE 802.3 [CES-12] and also IEEE 802.1Q [CES-11].

Another important path control mechanism that provides reliability is Frame replication and elimination. In this scenario, the IEEE 802.1CB Frame Replication and Elimination standard (FRER) [CES-13] defines how TSN frames belonging to a critical stream can be multiplied and sent through different paths towards their destination. This protects the stream against faults in any of the paths. The same mechanism ensures that the duplicates are merged and the excess is eliminated, to guarantee resource hygiene. At the point where the paths are joined and the extra-Frames eliminated, the redundancy ends.

3.3.2.2.2. Deterministic Networking (DetNet)

SDO: Internet Engineering Task Force (IETF).

DetNet is a networking technology that aims to provide determinism to the IP Layer 3. It delivers data flows with extremely low packet loss and bounded end-to-end delivery latency. This is possible with the active reservation of network resources, such as buffer space or transmission slots. [CES-14] In the current state of standardization, DetNet is set to operate on top of IP or Multiprotocol Label Switching (MPLS) setups. Additionally, since its conception, DetNet has been developed to interoperate with TSN, the other prominent deterministic initiative for wired networks. In the same way, integrations with 5G and other innovative wireless technologies are in the works.

The status of standardization of DetNet is shown in table 1. Despite its premature state, commercial solutions based on this technology have already been advertised.

Table 2. DetNet standard Status (as of October 21, 2022) (IETF)

Ready	Complete, under review	Ongoing standardization
RFC 8557 DetNet Problem Statement	draft-ietf-detnet-yang	draft-ietf-detnet-mpls-over-ip-preof
RFC 8578 DetNet Use Cases	draft-ietf-detnet-bounded-latency	draft-ietf-detnet-oam-framework
RFC 8655 DetNet Architecture		draft-ietf-detnet-mpls-oam
RFC 9055 DetNet Security Considerations		draft-ietf-detnet-ip-oam
RFC 8938 DetNet Data Plane Framework		draft-ietf-detnet-controller-plane-framework
RFC 8939 DetNet Data Plane: IP		draft-ietf-detnet-pof
RFC 8964 DetNet Data Plane: MPLS, RFC 9025 DetNet Data Plane: MPLS over UDP/IP, RFC 9056 DetNet Data Plane: IP over MPLS, RFC 9023 DetNet Data Plane: IP over TSN, RFC 9037 DetNet Data Plane: MPLS over TSN, RFC 9024 DetNet Data Plane: TSN VPN over MPLS, RFC 9016 DetNet Flow Information Model		

3.3.3. Review of the DATA-01-05 cluster

aerOS has been funded under the topic DATA-01-05-2021, together with other 5 Research and Innovation Actions. These five “sister” projects are targeting very similar goals as aerOS, covering from different perspectives the “demand” side or de “supply” side of meta operating systems for the continuum. In addition, some of them are tilted towards specific technical domains according to the structure proposed by the DG-CNECT of the European Commission (see next figure):

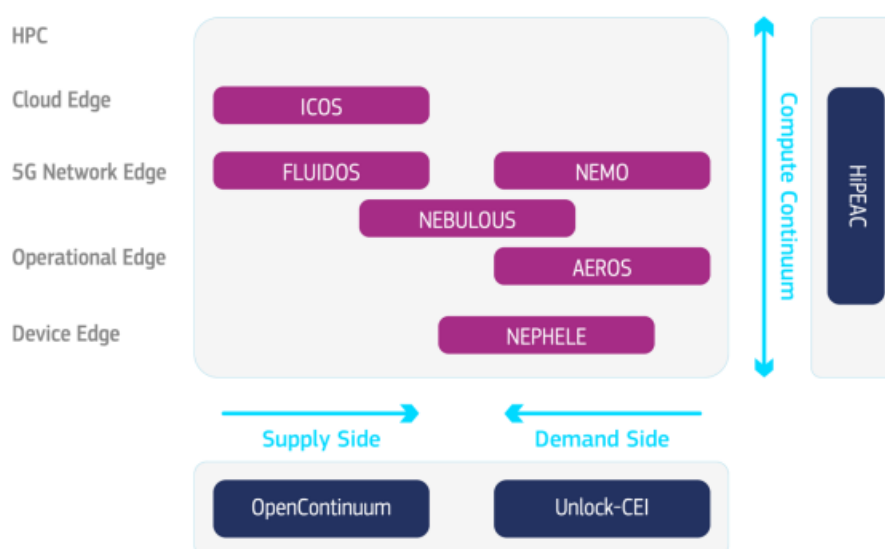


Figure 64: Meta-Operating Systems for the next generation IoT and Edge Computing - Source: Factsheet for Horizon Europe, Cluster 4, Destination 3: “Future European Platforms for the Edge: Meta-Operating Systems”

The projects are the following:

- FLUIDOS - Flexible, scaLable and secUre decentralISed Operationg System** ([Flexible, scaLable and secUre decentralISed Operations](#) | [FLUIDOS Project](#) | [Fact Sheet](#) | [HORIZON](#) | [CORDIS](#) | [European Commission \(europa.eu\)](#)): The idea behind a project like FluidDOS is born from the opportunity provided by the enormous processing capacity at the Edge, currently almost completely unused and spread across heterogeneous Edge devices that struggle to integrate with each other and to form a coherent computing continuum. The solution aims at being a disruptive, open-source paradigm that hinges upon secure protocols for advertisement and discovery, AI-powered resource orchestration and intent-based service integration. FluidDOS wants to create a fluid, dynamic, scalable and trustable computing continuum that spans across devices, unifies Edge and Cloud in a sustainable way, and possibly extends beyond administrative boundaries. Despite its innovation signature, FluidDOS will build upon already consolidated Operating Systems and orchestration solutions like Kubernetes, on top of which it will provide a new, enriched layer enacting resource sharing through advertisement and agreement procedures (in the horizontal dimension), and hierarchical aggregation of nodes, inspired by Inter-domain routing in the Internet (in the vertical dimension). Intent-based orchestration will leverage advanced AI Algorithms to optimise costs and energy usage in the continuum, promoting efficient usage of Edge resources. A Zero-Trust paradigm will allow FluidDOS to securely control and access geographically diverse resources, while Trusted Platform Modules will provide strong isolation and guarantee a safe deployment of applications and services. The aforementioned goals will be reached thanks to an open, collaborative ecosystem, whose creation will be focused on the development of a multi-stakeholder market of Edge services and applications, promoting European digital autonomy and sovereignty. Stakeholders from different fields will be involved in pilots and demonstrator for: Intelligent Energy, Agriculture and Logistics, which will challenge FluidDOS capabilities to adapt to different environments and operating conditions.

HORIZON EUROPE

- Grant agreement ID:** 101070473
 - Start date:** 1 September 2022 - **End date:** 31 August 2025
 - Funded under:** Digital, Industry and Space
 - Total cost:** € 8,415 433.95 - **EU Contribution:** € 8,406 433.95
 - Coordinated by:** MARTEL INNOVATE BV (Netherlands)
- ICOS - Towards a functional continuum operating system** ([Towards a functional continuum operating system](#) | [ICOS Project](#) | [Fact Sheet](#) | [HORIZON](#) | [CORDIS](#) | [European Commission \(europa.eu\)](#)): The proliferation of novel computing and sensing device technologies is a constantly increasing phenomenon of our time, and the growing demand for data-intensive applications in the Edge and Cloud, are driving a paradigm shift in computing around dynamic, intelligent and yet seamless interconnection of IoT, Edge and Cloud resources, in a single continuum. ICOS solution is intended to be an extended, open, secure, trustable, adaptable, technology agnostic and much more complete management strategy, covering the full continuum, e.g., IoT-to-Edge-to-Cloud, with a specific focus on the network connecting the whole stack, leveraging off-the-shell technologies (e.g., AI, data, and so on.), but also open to accommodate novel services as technology progress goes on. The ICOS project aims at proposing an approach embedding a well-defined set of functionalities, ending up in the definition of an IoT2cloud Operating System (ICOS). The main objective of the project is to conceptualise, to develop and to validate a Meta Operating System for a continuum, by facing unresolved matters such as devices heterogeneity, continuum infrastructure virtualisation and diverse network connectivity. ICOS intends to give an optimised and scalable service execution and performance and to guarantee trust, security and privacy, in addition to the reduction of integration costs and effective mitigation of Cloud provider lock-in effects, in a data-driven system built upon the principles of openness, adaptability and data sharing.

HORIZON EUROPE

- Grant agreement ID:** 101070177

- **Start date:** 1 September 2022 - **End date:** 31 August 2025
 - **Funded under:** Digital, Industry and Space
 - **Total cost:** € 10,997.675 - **EU Contribution:** € 10,997.675
 - **Coordinated by:** ATOS SPAIN SA (Spain)
- **NEBULOUS - A Meta Operating System For Brokering Hyper-distributed Applications On Cloud Computing Continuums** ([A META OPERATING SYSTEM FOR BROKERING HYPER-DISTRIBUTED APPLICATIONS ON CLOUD COMPUTING CONTINUUMS](#) | [NebulOus Project](#) | [Fact Sheet](#) | [HORIZON](#) | [CORDIS](#) | [European Commission \(europa.eu\)](#)): Cloud computing is a centralised system. Fog computing is a distributed decentralised infrastructure that bridges the gap between the cloud and IoT devices. In the realms of cloud and fog computing brokerage, it's important to introduce advanced methods and tools. This is the aim of the EU-funded NebulOus project. NebulOus will enable secure and optimal application provisioning and reconfiguration over the cloud computing continuum. Specifically, it will develop a novel Meta Operating System and platform for enabling transient fog brokerage ecosystems that seamlessly exploit edge and fog nodes. This will be in conjunction with multi-cloud resources, to cope with the requirements posed by low latency applications.
NebulOus will accomplish substantial research contributions in the realms of cloud and fog computing brokerage by introducing advanced methods and tools for enabling secure and optimal application provisioning and reconfiguration over the cloud computing continuum. NebulOus will develop a novel Meta Operating System and platform for enabling transient fog brokerage ecosystems that seamlessly exploit edge and fog nodes, in conjunction with multi-cloud resources, to cope with the requirements posed by low latency applications. The envisaged BRONCO solution includes the following main directions of work:
I) Development of appropriate modelling methods and tools for describing the cloud computing continuum, application requirements, and data streams; these methods and tools will be used for assuring the QoS of the provisioned brokered services.
II) Efficient comparison of available offerings, using appropriate multi-criteria decision-making methods that are able to consider all dimensions of consumer requirements.
III) Intelligent applications, workflows and data streams management in the cloud computing continuum.
IV) Addressing in a unified manner the security aspects emerging in of transient cloud computing continuums (e.g., access control, secure network overlay etc.).
V) Conducting and monitoring smart contracts-based service level agreements.

HORIZON EUROPE

- **Grant agreement ID:** 101070516
 - **Start date:** 1 September 2022 - **End date:** 31 August 2025
 - **Funded under:** Digital, Industry and Space
 - **Total cost:** € 8,478 106.25 - **EU Contribution:** € 8,478 106.25
 - **Coordinated by:** FUNDACIO EURECAT (Spain)
- **NEMO – Next Generation Meta Operating system** ([Next Generation Meta Operating System](#) | [NEMO Project](#) | [Fact Sheet](#) | [HORIZON](#) | [CORDIS](#) | [European Commission \(europa.eu\)](#)): NEMO established as a main goal the introduction of an open source, flexible, adaptable, cybersecure and multi-technology Meta Operating System, sustainable during and after the end of the project, via the Eclipse foundation (NEMO consortium member). To achieve technology maturity, NEMO will take existing systems as a starting point, together with technologies and Open Standards, while introducing novel concepts, tools, Living Labs and engagement campaigns to go beyond the state of the art.

NEMO will introduce innovations at different layers of the protocol stack, enabling on-device Cybersecure Federated ML/DRL, deliver time-triggered (TSN) multipath ad-hoc/hybrid self-organised and zero-delay failback/self-healing multi-cloud clusters, multi-technology Secure Execution Environment and on-Service Level Objectives Meta-Orchestrator, Plugin and Apps Lifecycle Management and Intent Based programming tools. Furthermore, NEMO will be cybersecure and trusted adopting Mutual TLS and Digital Identity Attestation.

The solution will be validated through 5 pilots in the following industrial sectors: Farming, Energy, Mobility/City, Industry 4.0 and Media/XR. In addition, 8 use cases in Living Labs, using more than 30 heterogeneous IoT devices and real 5G infrastructure. The impact will not only safeguard EU position in data economy and applications verticals, but lower energy efficiency, reduce pesticides and Carbon Footprint.

HORIZON EUROPE

- **Grant agreement ID:** 101070118
 - **Start date:** 1 September 2022 - **End date:** 31 August 2025
 - **Funded under:** Digital, Industry and Space
 - **Total cost:** € 10,499.650 - **EU Contribution:** € 10,499.650
 - **Coordinated by:** ATOS SPAIN SA (Spain)
- **NEPHELE - A Lightweight Software Stack and Synergetic Meta-orchestration Framework For The Next Generation Compute Continuum ([A LIGHTWEIGHT SOFTWARE STACK AND SYNERGETIC META-ORCHESTRATION FRAMEWORK FOR THE NEXT GENERATION COMPUTE CONTINUUM | NEPHELE Project | Fact Sheet | HORIZON | CORDIS | European Commission \(europa.eu\)](#)):** The project NEPHELE aims at the management of reliable and secure end-to-end hyper-distributed applications across heterogeneous infrastructure in the Cloud-to-Edge-to-IoT continuum, the convergence of IoT technologies and the development of synergetic orchestration mechanisms. Several use cases across various vertical industries are considered by the project, including Disaster Management, Logistic Operations in Ports, Energy Management in Smart Buildings and Remote Healthcare services. Two successive open calls will also take place, while a wide open-source community is envisaged to be created for supporting the intended outcomes.

The scope of NEPHELE is to use programmable infrastructure that is spanning across the compute continuum from Cloud-to-Edge-to-IoT, removing existing openness and interoperability barriers in the convergence of IoT technologies against Cloud and Edge computing orchestration platforms, and introducing automation and decentralised intelligence mechanisms powered by 5G and distributed AI technologies.

The NEPHELE project aims to introduce two core innovations:

- I) an IoT and Edge computing software stack for virtualisation of IoT devices at the Edge part of the infrastructure and supporting openness and interoperability aspects in a device-independent way.
- II) a synergetic meta-orchestration framework for managing the coordination between Cloud and Edge computing orchestration platforms, through high-level scheduling supervision and definition.

HORIZON EUROPE

- **Grant agreement ID:** 101070487
- **Start date:** 1 September 2022 - **End date:** 31 August 2025
- **Funded under:** Digital, Industry and Space
- **Total cost:** € 9,127 711.25 - **European Contribution:** € 9,127 711.25
- **Coordinated by:** ETHNICON METSOVION POLYTECHNION (Greece)

3.3.4. Other related projects

In this chapter, the global status of research of the computing continuum is overviewed. In addition, a series of other relevant projects to (not in the same clusters of) aerOS are analysed. A few details are provided in order to understand potential synergies, differences and similarities in scope.

3.3.4.1. Global Analysis of the European Research on the Edge-Cloud Computing Continuum

For a decisive push to an efficient long-term realisation of such a significant field as the IoT Edge-Cloud Continuum, the following key principles have been identified by the main lines of research through the European Union:

1. The growth in computing capabilities for smart devices (e.g., tiny edges): novel devices have enough resources to run applications with substantial and ever-increasing complexity, security, privacy and trust. This opens a new potential in the level of distribution and granularity of the computation resources that the IoT edge-cloud continuum can utilise.
2. The maturity of the multi-domain orchestration tools: with regard to virtualised and containerised functions, the computation power management has evolved providing a rich toolset, based on related releases, such as NFV Release 4, or powerful specialised cloud infrastructure software stacks for the edge, e.g., StarlingX, OneEdge. This provides the guarantees for a flexibly and fully-orchestrated virtualisation and containerisation-based environment.
3. The enabling of programmability for the edge segments: the recent Cloud Industry Roadmap, together with standardisation on exposure capabilities for access network domains (ETSI GS MEC 009), opens a clear and decisive window for application-driven monitoring and control of resources (meaning storage, compute, network) at any domain within the path between constrained devices and cloud. This allows third-party developers to tightly integrate applications to the network infrastructure.
4. The Artificial Intelligence potential: the concept of open-source has allowed the rise of data-based intelligence, above all because of the vastness of data that is becoming accessible, with AI as the game-changer, in the decision making within the IoT edge-cloud continuum.

The IoT ecosystem is a dynamic aggregation of resources, e.g., sensors, actuators, processing and storage, populating edges of current infrastructures, e.g., edge computing with local ad-hoc clouds, fog computing, far edge and federated approaches. Artificial Intelligence and real-time processing may require high computing power close to events and, sometimes, distributed across Infrastructure Elements. Horizon 2020 and, in general, European Union funded projects, like ACCORDION and DECENTER (described, among others, in the general overview of section 3.3.4.1.), already address continuum challenges, by associating edge computing with 5G, and by realising Fog Computing platform. This distributed data and compute scenario is called Network Compute Fabric [ORP-1]: in a context like that, the network should host computing intertwined with communication for the highest level of efficiency, in order to properly support heterogeneous systems, that range from simple terminals to performance-sensitive robots and Augmented Reality (AR) nodes. It must be noted that Edge Meta-Operating Systems absolutely require flexibility to serve any possible dynamic combination of infrastructure elements while providing globally orchestrated services (e.g., policy services specifying behaviour; data governance; or even cognitive services) [ORP-2].

Regarding the State of the Art for Edge Meta Operating Systems, the following developments must be underlined: Thin Edge, ROS for robotic environments, EOS [ORP-3] for virtualised telco networks, or VirtuOS [ORP-4] for the cloud. IoT edge-cloud continuum orchestration Service orchestration follows recent advances in SDN/NFV, e.g., Cloud-Native functions (i.e., CNFs). Orchestration provides seamless, elastic service deployment for verticals, while efficiently reusing the available resources, reducing incurred costs and consumed energy. A challenge that has to be faced is the need to properly orchestrate services in a heterogeneous continuum of resource federation, in opposition to single-domain orchestration where the orchestrator has full control over resources. A multi-domain orchestration, instead, requires coordination across domains [ORP-5]. There are some alternative options in terms of centralised, distributed and hierarchical orchestrators, in which the growing complexity, calls for automated orchestration and management of services [ORP-6]. Different initiatives exist, like ETSI ZSM ISG; ETSI ENI ISG; ZOOM by TMF, Open RAN, NWDAF (Network Data

Analytics Function), ETSI OSM, ETSI MEC ISG [ORP-7]. Network and service providers build their business logic around microservices and AI. Orchestrators map high-level QoS requirements into appropriate set of tasks characterised by resource requirements, their locations, and level of isolation. Currently, resource allocations to network components are handcrafted by the operators, leading to resources over(under)provisioning. Therefore, data and event-driven service orchestration is needed to allocate the right number of resources to each slice [ORP-8].

IoT edge-cloud continuum smart networking service deployment and reconfiguration across IoT edge-cloud continuum is challenging mostly because of the heterogeneity of the network. Standalone services have to face network requirements concerning data sources, to be fulfilled by leveraging technologies related to NFV and SDN, but also 5G Network Programmability via the native service APIs (3GPP NEF/SEAL/CAPIF) and the 3GPP vertical application enablers, such as the EDGE_APP.

Composition of services with heterogeneous requirements (e.g., latency) [ORP-9] can also be enacted vertically, where reconfiguration of services (and network, if necessary) is even more complex. Furthermore, devices are increasingly becoming smarter in collecting, processing and transmitting data, while the incredible growth of connected devices and sensors is promoting novel, computationally intensive, IoT applications that can cause network bottlenecks, impacting overall performance. Therefore, it is mandatory to apply new techniques in order to provide better support for IoT operations across IoT edge-cloud continuum, while at the same time preventing any unnecessary communication that might affect the performance of the network, and reducing costs of data storage and computation. Networks are key to achieve increasingly demanding levels of reconfigurability and automation, in order to scale efficiently, manage resources, and optimise operation while handling multi-vertical traffic with distinct demands [ORP-10].

IoT ecosystems are comprised of heterogeneous multi-vendor nodes, thus creating a huge discrepancy in their capabilities and resources (e.g., processing power or storage capacity), and their underlying hardware. Virtualisation allows services and applications to run in a homogeneous environment, no matter the hardware or operating system. Standardised APIs allow services to access specific hardware e.g., GPUs, memory or storage [ORP-11]. Moreover, clustering multiple virtualised nodes delivers large federated pool of resources. To meet the instance of allowing adequate resource continuity, the compute continuum architecture needs a common infrastructure virtualisation framework. Although VMs are common for the cloud, they are not suitable for constrained devices and edge nodes, because of the large overhead they add. For optimal resource allocation and high QoS, virtualisation frameworks should be tailor-made for each specific domain with its specific requirements [ORP-12], while being entirely hardware-independent. Different frameworks strive at achieving this goal: Docker Swarm, Kubernetes, FITOR [ORP-13], EPOS Fog [ORP-14], Apache Mesos, and several others, are already well established in the cloud but they have to be adapted to the heterogeneous nature of the IoT edge-cloud continuum distributed and federated deployments, to provide scalable continuum of resources [ORP-15].

When speaking about a crucial topic such as data sovereignty, it is facing the combined ability to keep data within a particular realm, and the explicit knowledge and control on how data are processed, stored, and forwarded. Furthermore, data autonomy is related to the capacity of homogenising data models at the edge, e.g., to query, interoperate or prepare data to be used by AI modules. Current practices in data processing are focused on access control and enforcement of secure forwarding and storage, with different identity schemas (for example, centralised, distributed, federated), authorisation models and access policies [ORP-16].

Intensive use of data evidence for control and management processes needs:

- Usability: data are provided according to the structure required by consumers.
- Sufficiency: data are generated by required sources and processors, according to a planned topology.
- Safety: data provenance related properties (e.g., origin, timeframe) can be verified.
- Steadiness: availability and continuity of data flows are assured. Most, if not all, of these properties are associated with availability of well-structured and sufficient metadata [ORP-17] to manage data access, forwarding and processing.

According to RAMI4.0, AI can be beneficial not just at functional but also at business level (e.g., IEEE Ethically Aligned Design for Business), when concerns about its reliability and safety are addressed. AI may support an efficient decision-making, e.g., optimise sequencing of activities that run at different IoT/edge nodes, and/or the

cloud (referring to critical operations, such as those found in aerOS use cases: forecasts/planning in logistics, production, downtimes, resource availability, etc.) [ORP-18]. Edge resource constraints bring challenges, but frugal AI methods may provide solutions. While frugal AI approaches are a hot research topic, they are studied using “cloud resources”. Besides, AI clearness may be needed in the real-world, requiring additional resources and overcoming problems caused by streaming data, so it is also pursued (mostly) in the cloud. Separately, IoT/edge ecosystems naturally match federated/distributed AI/ML scenarios [ORP-19]. However, existing frameworks still have constraints to address. Meta operating systems support for cybersecurity is a multi-dimensional problem of protection of data stored, in transit, and during processing.

Increase in security needs, raised by processing data locally, causes novel challenges to be addressed [ORP-20]:

- requirements for lightweight data encryption and fine-grained data sharing;
- heterogeneous data dissemination control and secure data management;
- balancing security between large-scale edge services and resource-constrained edge devices;
- efficient privacy preserving mechanisms.

It must be noticed that data governance is already a challenge on her own, considering how data are scattered across several levels and thus need to be stored, deleted, processed, searched, transmitted and accessed [ORP-21] while keeping security, integrity, trust and privacy [ORP-22]. The inherent distributed nature of IoT edge-cloud continuum, poses security and privacy challenges due to the heterogeneity of edge infrastructural elements and migration of services among them. A potential solution could be based on DLT, providing reliable access and control of the network, enhancing data integrity and computation validity [ORP-23]. Furthermore, research challenges have to be addressed in terms of security, privacy and trust, with focus on scalability [ORP-24], and the extension of DevSecOps methodology to include privacy by design.

3.3.4.2. Horizon Europe and Horizon 2020 projects with similar goals

Other relevant projects that are being continuously monitored due to their relevance for aerOS goals are:

- **SERRANO - Transparent Application Deployment In A Secure, Accelerated And Cognitive Cloud Continuum (TRANSPARENT APPLICATION DEPLOYMENT IN A SECURE, ACCELERATED AND COGNITIVE CLOUD CONTINUUM | SERRANO Project | Fact Sheet | H2020 | CORDIS | European Commission (europa.eu))**: The SERRANO project aims at introducing a novel ecosystem of Cloud-based technologies, with an abstraction layer that transforms the distributed Edge, Cloud and high-performance computing resources into a single borderless infrastructure, thus simplifying their automated and cognitive orchestration. These aspects will enable application-specific service instantiation and optimal customisations based on the workloads to be processed, in a holistic manner, thus supporting highly demanding, dynamic and security-critical applications.

SERRANO is tuned and completely aligned with current trends in the Cloud computing sector towards the extension of Cloud infrastructures in order to properly integrate Edge resources. It proposes the introduction and evolution of novel key concepts and approaches that aim at resolving existing technology gaps, towards the establishment of advanced infrastructures, able to meet the stringent requirements of future applications and services. It will develop technologies and mechanisms related to security and privacy in distributed computing and storage infrastructures, hardware and software acceleration on Cloud and Edge, cognitive resource orchestration, dynamic data movement and task offloading between edge/cloud/HPC, transparent application deployment, energy-efficiency and real-time and zero-touch adaptability. SERRANO will demonstrate its solution through three use cases related to: secure Cloud and Edge storage over a diversity of Cloud resources; fintech by supporting latency-sensitive and safety-critical digital services in the financial sector; machine anomaly detection in manufacturing for Industry 4.0.

H2020

- **Grant agreement ID:** 101017168
- **Start date:** 1 January 2021 - **End date:** 31 December 2023
- **Funded under:** INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

- **Total cost:** € 4,343.180 - **European Contribution:** € 4,343.180
- **Coordinated by:** INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS (Greece)
- **ACCORDION - Adaptive edge/cloud compute and network continuum over a heterogeneous sparse edge infrastructure to support nextgen applications** ([Adaptive edge/cloud compute and network continuum over a heterogeneous sparse edge infrastructure to support nextgen applications | ACCORDION Project | Fact Sheet | H2020 | CORDIS | European Commission \(europa.eu\)](#)): The ACCORDION project aims at associating Edge computing with advanced technologies such as 5G, so that the EU will be able to capitalise on its local resource and infrastructure and bring benefit to the SMEs throughout its territory. The project uses a practical approach in connecting edge resources and infrastructures to support next-generation applications. Considering that Edge computing is intrinsically more “democratic” than Cloud computing., the idea to synergistically employ Edge computing with upcoming technologies such as 5G provides a great opportunity for EU to capitalise on its local resource and infrastructure and its SME-dominated application development landscape and achieve an Edge-computing-driven disruption with a local business scope. Therefore, ACCORDION tries to bring together Edge infrastructures (public Clouds, on-premise infrastructures, telco resources, even end-devices) in pools defined in terms of latency, that can support NextGen application requirements. It will also intelligently orchestrate the compute and network continuum formed between Edge and public Clouds, using the latter as a capacitor. Deployment decisions will be taken also based on privacy, security, cost, time and resource type criteria. The slow adoption rate of novel technological concepts from the EU SMEs will be tackled through an application framework, that will leverage DevOps and SecOps to facilitate the transition to the ACCORDION system. With a strong emphasis on European edge computing efforts (MEC, OSM) and 3 highly anticipated NextGen applications on collaborative VR, multiplayer mobile- and cloud-gaming, brought by the involved end users, ACCORDION is expecting to radically impact the application development and deployment landscape, also directing part of the related revenue from non-EU vendors to EU-local infrastructure and application providers.

H2020

- **Grant agreement ID:** 871793
- **Start date:** 1 January 2020 - **End date:** 31 December 2022
- **Funded under:** INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)
- **Total cost:** € 4,754 738.75 - **European Contribution:** € 4,754 738.75
- **Coordinated by:** CONSIGLIO NAZIONALE DELLE RICERCHE (Italy)
- **DECENTER - Decentralised technologies for orchestrated cloud-to-edge intelligence** ([Decentralised technologies for orchestrated cloud-to-edge intelligence | DECENTER Project | Fact Sheet | H2020 | CORDIS | European Commission \(europa.eu\)](#)): The DECENTER project wanted to create a solid fog computing platform offering AI application-aware orchestration and provisioning of resources accommodating cross-border collaboration between cloud and IoT providers.

The analysis of the state of the art made by the consortium showed that AI required high computational resources only available in high-performance data centres; therefore, realising an architecture capable of securely processing this unprecedented amount of remotely sensed and potentially sensitive data, as well as conveying timely responses to pervasive configurable actuators was a significant endeavour. The project tried to improve existing Cloud and IoT solutions with advanced capabilities to abstract features and process data closer to where it is produced, while enabling a collaborative environment in which multiple stakeholders (Cloud and IoT providers) was securely able to share and harmoniously manage resources, in dynamically created multi-Cloud/Edge, federated environments. Cross-border infrastructure federation would be realised via Blockchain-based Smart Contracts defining customised Service Level Agreements, used to commit the execution of verified workloads across multiple, potentially remote, administrative domains. Thus, DECENTER would unlock the potential of

innovative decentralised AI algorithms and models, by deploying them across multiple tiers of the infrastructure and federated clouds.

The project validated its solution with real-world pilots executed in urban, industrial and home environments. With its approach, DECENTER targeted the emergence of innovative digital businesses, thus providing a competitive advantage to EU and Korean industry and fostering cross-border collaboration.

H2020

- **Grant agreement ID:** 815141
- **Start date:** 1 July 2018 - **End date:** 30 June 2021
- **Funded under:** INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)
- **Total cost:** € 2,197.700 - **European Contribution:** € 2,197.700
- **Coordinated by:** FONDAZIONE BRUNO KESSLER (Italy)
- **LIGHTKONE - Lightweight Computation for Networks at the Edge** ([Lightweight Computation for Networks at the Edge | LightKone Project | Fact Sheet | H2020 | CORDIS | European Commission \(europa.eu\)](#)): The main goal of a project like LightKone was the development of a scientifically sound and industrially validated model for doing general-purpose computation on Edge networks, which consist of a large set of heterogeneous, loosely coupled computing nodes situated at the logical extreme of a network. Well-known examples are networks of Internet of Things, mobile devices, personal computers, and points of presence including Mobile Edge Computing. When the project was designed, internet applications were already increasingly running on Edge networks in order to reduce latency, increase scalability, resilience, and security, and permit local decision making. Despite this, the current market did not provide any solution for the definition of general-purpose computations on Edge networks, e.g., computation with shared mutable state. LightKone tried to solve this problem by combining two recent advances in distributed computing, namely synchronisation-free programming and hybrid gossip algorithms, both of which were, and still are, successfully used separately in industry. Together, they formed a natural combination for Edge computing. The intention was to cover Edge networks both with and without data centre nodes, and applications focused on collaboration and computation, separately and combined. Project results were intended to be new programming models and algorithms that advance scientific understanding, implemented in new industrial applications and a start-up company, and evaluated in large-scale realistic settings.p

H2020

- **Grant agreement ID:** 732505
- **Start date:** 1 January 2017 - **End date:** 31 December 2019
- **Funded under:** INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)
- **Total cost:** € 3,570, 993.75 - **European Contribution:** € 3,570, 993.75
- **Coordinated by:** UNIVERSITE CATHOLIQUE DE LOUVAIN (Belgium)
- **PRESTOCLOUD - PrEstoCloud - Proactive Cloud Resources Management at the Edge for Efficient Real-Time Big Data Processing** ([PrEstoCloud - Proactive Cloud Resources Management at the Edge for Efficient Real-Time Big Data Processing | PrEstoCloud Project | Fact Sheet | H2020 | CORDIS | European Commission \(europa.eu\)](#)): PrEstoCloud project made substantial research contributions in the Cloud computing and real-time data intensive applications domains, in order to provide a dynamic, distributed, self-adaptive and proactively configurable architecture for processing Big Data streams. In particular, PrEstoCloud aimed at combining real-time Big Data, mobile processing with Cloud computing research in a unique way that wanted to entail proactiveness of Cloud resources use and expansion of the Fog computing paradigm to the extreme Edge of the network. The envisioned

solution was driven by the microservices paradigm and has been structured across five different conceptual layers: Meta-management; Control; Cloud infrastructure; Cloud-Edge communication and Devices layers. The innovative character of the solution was tested through three PrEstoCloud pilots from the Logistics, Mobile journalism and Security surveillance application domains.

H2020

- **Grant agreement ID:** 732339
- **Start date:** 1 January 2017 - **End date:** 31 December 2019
- **Funded under:** INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)
- **Total cost:** € 4,256 502.50 - **European Contribution:** € 4,256 502.50
- **Coordinated by:** SOFTWARE AG (Germany)

From the perspective of 3GPP and 5GPP, the following considerations must be provided. 5G technological and architectural features that will shape the new access, networking, and management domains in mobile communications are being developed and tested across Europe. These features promise countless opportunities for service innovation and business efficiencies, creating an unprecedented impact on multiple vertical sectors². The first wave of 5G standards (3GPP Release 15) has been released, while, many cutting-edge technologies, resulting from huge private and public research investment within the industry and a series of 5G-PPP projects³, are pushing their way towards higher technology readiness levels (TRL) and eventual commercialization. The next 5G release is focused on industrial applications and involves multiple trials across 28 member states, conducting both conforming and field trials for concurrent support of heterogeneous 5G use cases set by multiple vertical sectors, including the five major vertical sectors defined by 5G-PPP, namely Media & Entertainment, Public Protection and Disaster Relief (PPDR), e-Health, Automotive, and Industry 4.0.

5G vertical trials in Europe have been performed through 5G Public Private Partnership projects (5G-PPP) funded by 700M€ of the European Union research funding grants and matched by 3,5B€ of private funding in the 2014-2020 timeframe. The 5G Infrastructure Public Private Partnership (5G-PPP) is a joint initiative between the European Commission and European ICT industry (ICT manufacturers, telecommunications operators, service providers, SMEs and researcher Institutions). The 5G-PPP is now in its third phase where many new projects were launched in Brussels initially in June 2018 and more followed in 2019 and 2020. The 5G-PPP will deliver solutions, architectures, technologies, and standards for the ubiquitous next generation communication infrastructures of the coming decade. The challenge for the 5G Public Private Partnership (5G-PPP) is to secure Europe's leadership in the particular areas where Europe is strong or where there is potential for creating new markets such as smart cities, e-health, intelligent transport, education, or entertainment and media⁴.

The underlying technology developed in the context of the 5G-PPP Initiative was a key enabler for many success stories. The 5G-PPP Initiative has provided a number of scientific solutions that have been contributed to standardization activities and also the global academic and research community through publications. In addition, the 5G-PPP projects have been driving test and validation activities in Europe, collecting significant experience for all stakeholders, and raising public awareness on the capabilities of 5G networks. The whole 5G-PPP trial project portfolio is now worth more than EUR 300 million of EU funding and is expected to leverage more than EUR 1 billion of private investment in 5G vertical trials, reinforcing Europe's leading position in this field⁵.

As the last project calls for H2020/5G-PPP took place, it is worth pointing out that the development of mobile communication technology will not stop with the end of this Programme. The last 5G-PPP project calls will be the first set of projects to consider what comes after 5G. These Beyond 5G (B5G) projects should provide the

² Vertical sectors: <https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf>

³ 5G-PPP projects <https://5g-ppp.eu/5g-ppp-phase-3-projects/>

⁴ <https://5g-ppp.eu/>

⁵ Full-5G-Annual-Journal-2020

bridge to the future activities foreseen in the next Smart Networks and Services (SN&S) partnership Programme which is proposed to be part of Horizon Europe.

5G-PPP Phases and ICT calls

More than half a decade after the launch of the 5G-PPP, first commercial 5G services are now available in a number of European cities and many 5G-PPP research projects are still ongoing. The 5G-PPP Initiative is organized in 3 different main Phases.

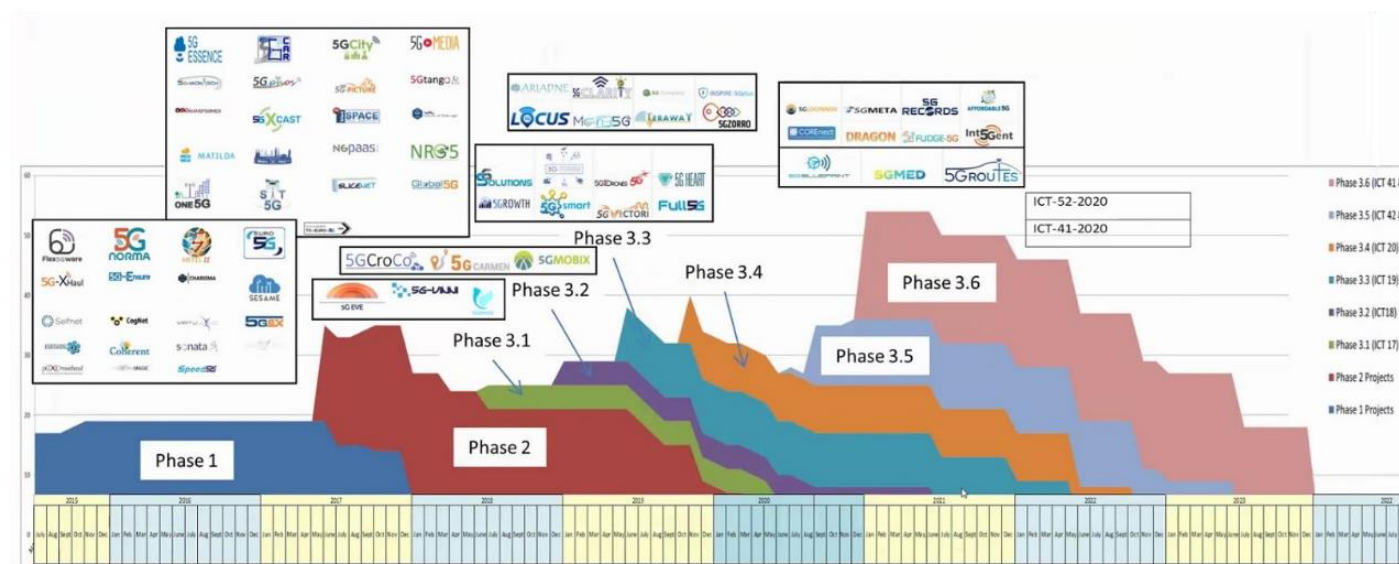


Figure 65. Overview of the 5G-PPP Programme⁶

The first phase (Phase 1) focused on basic research to provide the key concepts and solutions for 5G networks. The second phase (Phase 2) concentrated on bringing this new 5G technology to the vertical industries and finally Phase 3 where large-scale trials and innovation infrastructures are being created. The third phase (Phase 3) also contains basic research activities to consider evolution beyond 5G.

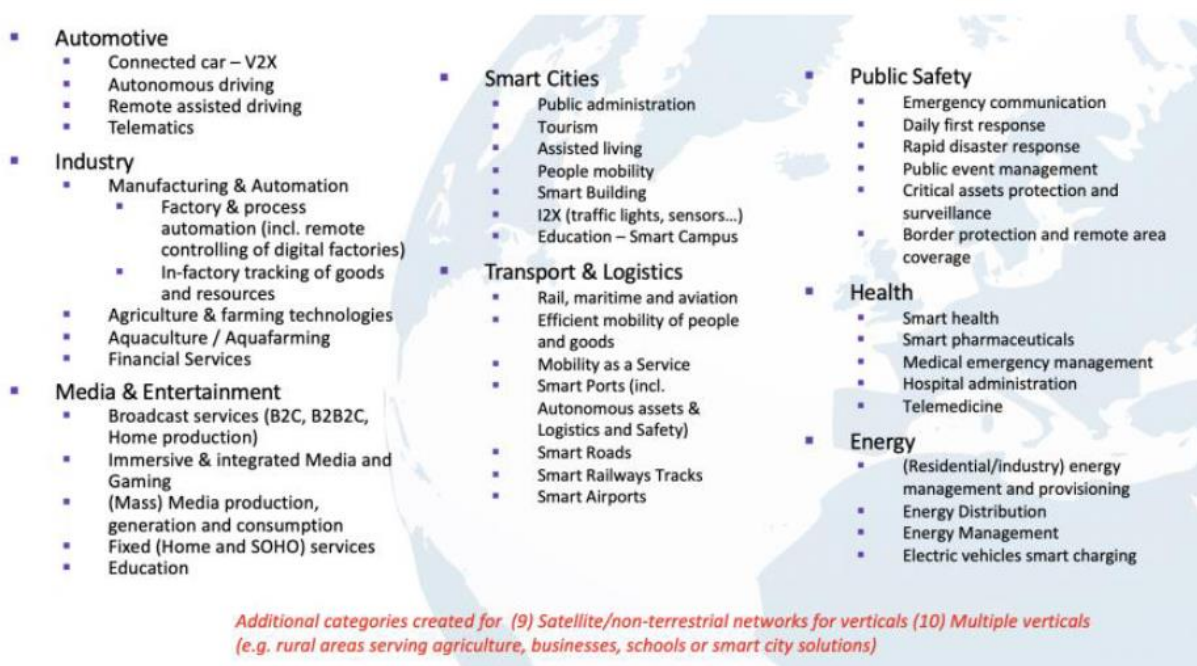


Figure 66 Mapping of use cases to vertical categories⁷

⁶ 5G-PPP Progress Monitoring Report <https://5g-ppp.eu/wp-content/uploads/2020/10/5G-PPP-PMR2019v1-6.pdf>

⁷ <https://global5g.org/>

The last two Phases of 5G-PPP have managed to cover a significant number of vertical industries as shown in Figure 66. This is an important achievement because one of the main aims of 5G is the support of the so call verticals. Phase 2 started in June 2017, with 21 new 5G-PPP projects, including 2 complementary CSA projects. These projects relied on the technologies, produced during Phase 1, for the digitization and integration of vertical industries in Europe. Most Phase 2 projects successfully completed in 2019, while some were continuing in 2020. This phase was more focused on demonstrating and validating the developed technology and explicitly trying to integrate use cases from vertical industries beyond classical tele-communications.

During 2018, the Phase 3 of the 5G-PPP framework was initiated with the first three Phase 3 projects. This involved essentially the roll out of 5G platforms across Europe. The target was to enable large scale trials to help the stakeholders testing, in realistic environments, the key findings from the previous phases and draw significant conclusions. In 2018, three infrastructure projects (ICT-17) were selected to create a pan-European large-scale 5G test platform to be used by a number of vertical use cases. During 2019, these projects have setup a significant part of their platforms and provided a clear and detailed roadmap of their features that will be offered in multiple sites all over Europe⁸ (refer to Figure 67, which presents the 5G Infrastructure PPP Phase 3 Platforms Projects – Geographic Cartography). Also, these projects have clearly identified how their platforms can be used for advanced testing by other 5G-PPP and not only research projects⁹.

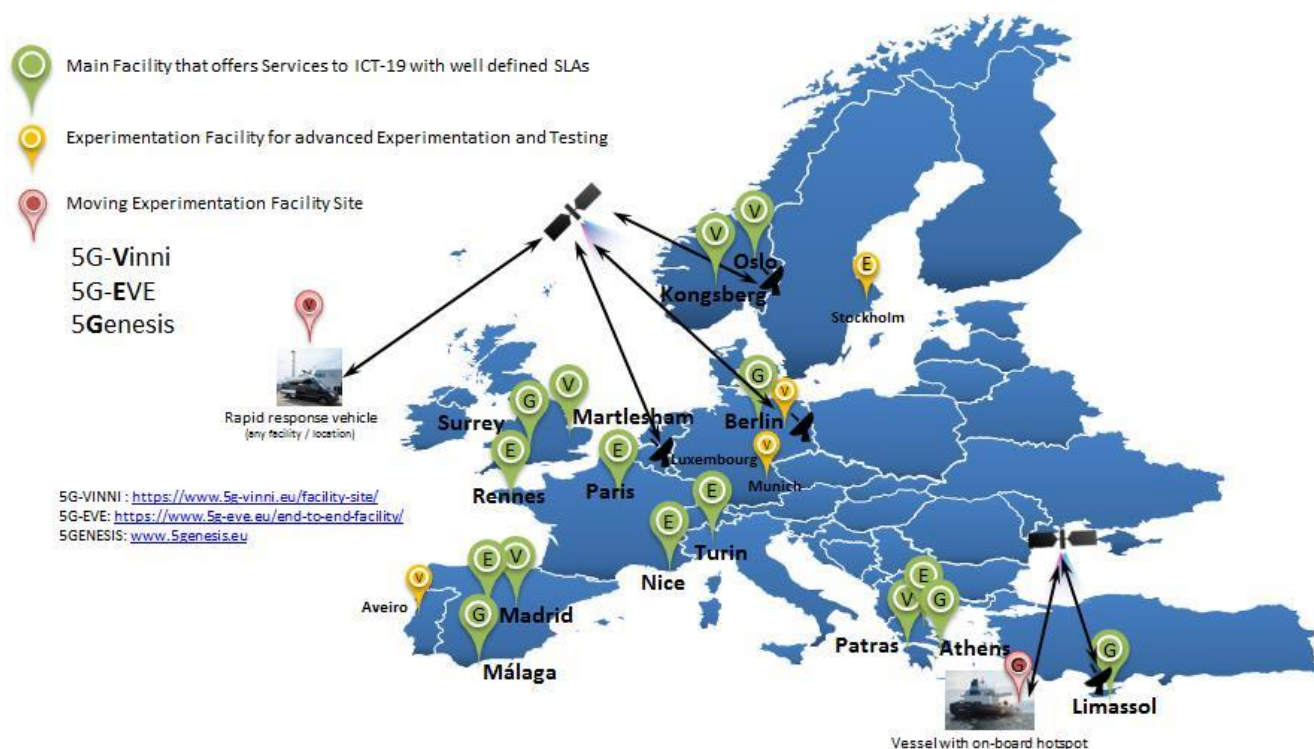


Figure 67. 5G Infrastructure PPP Phase 3 Platforms Projects – Geographic Cartography¹⁰

Also, in November 2018, three ICT-18 automotive/corridors projects started their activities implementing and testing advanced scenarios and one additional automotive project is also active in the context of EU-China Collaboration. During 2019 these projects have completed the identification of the use case to be validated in cross border/Mobile Network Operators/Vendor/Generation trials. They have identified network requirements, potential changes in the network architecture and provided recommendations for regulation and spectrum.

In relation to the ICT-19 projects (commenced June 2019), eight projects (seven R&I and one CSA projects) have been selected out from the 32 proposals that were evaluated by the EC, in response to the 5G-PPP ICT-

⁸ Technology Board white paper, 5G network support of vertical industries in the 5G-PPP ecosystem, February 2020, https://5g-ppp.eu/wp-content/uploads/2020/03/5PPP_VTF_brochure_v2.1.pdf

⁹ Technology Board white paper, On board procedure to 5G-PPP Infrastructure Projects, April 2020, <https://5g-ppp.eu/wp-content/uploads/2020/04/On-Board-Procedure-to-5G-PPP-Infrastructure-Projects-1.pdf>

¹⁰ <https://5g-ppp.eu/5g-ppp-platforms-cartography/>

19-2019 call. The projects mainly rely for their trials on the three ICT-17 platform projects, although some of them are also developing their own platforms to perform further testing.

The ICT-17 and ICT-19 projects are covering a significant number of vertical industries as shown in Figure 68. The first three rows illustrate the vertical industries being covered by the 3 ICT-17 projects while the remaining seven, present those covered by the ICT-19 projects.

	 Industry 4.0	 Agriculture & agri-food	 Automotive	 Transport & logistics	 Smart Cities & utilities	 Public Safety	 Smart (air)ports	 Energy	 eHealth & wellness	 Media & entertain.
5G EVE	✓		✓		✓	✓		✓	✓	✓
5GENESIS				✓	✓	✓				✓
5G VINNI	✓			✓		✓		✓		
5GIDRONES				✓		✓				✓
5G HEART		✓	✓	✓					✓	
5G GROWTH	✓			✓				✓		
5G SMART	✓									
5G SOLUTIONS	✓				✓		✓	✓		✓
5G TOURS				✓	✓		✓		✓	✓
5G VICTORI	✓			✓				✓		✓

Figure 68. Vertical industries under validation by ICT-17 and ICT-19 projects¹¹

In November 2019, and under the ICT-20 call, eight new projects have started working on the longer-term vision for telecommunication networks. These projects target providing innovative solutions to transform the network into a low energy distributed computer.

In such a system, processes and applications will be dynamically created, moved, and suppressed, depending on the information flows and customer needs. In the evolved networks, new terminal types based on gestures, facial expressions, sound, and haptics may also form the basis of the interaction between humans and infosystems. Figure 69 is the main Phase 3 reference figure of 5G-PPP.

¹¹ 5G-PPP Progress Monitoring Report <https://5g-ppp.eu/wp-content/uploads/2020/10/5G-PPP-PMR2019v1-6.pdf>

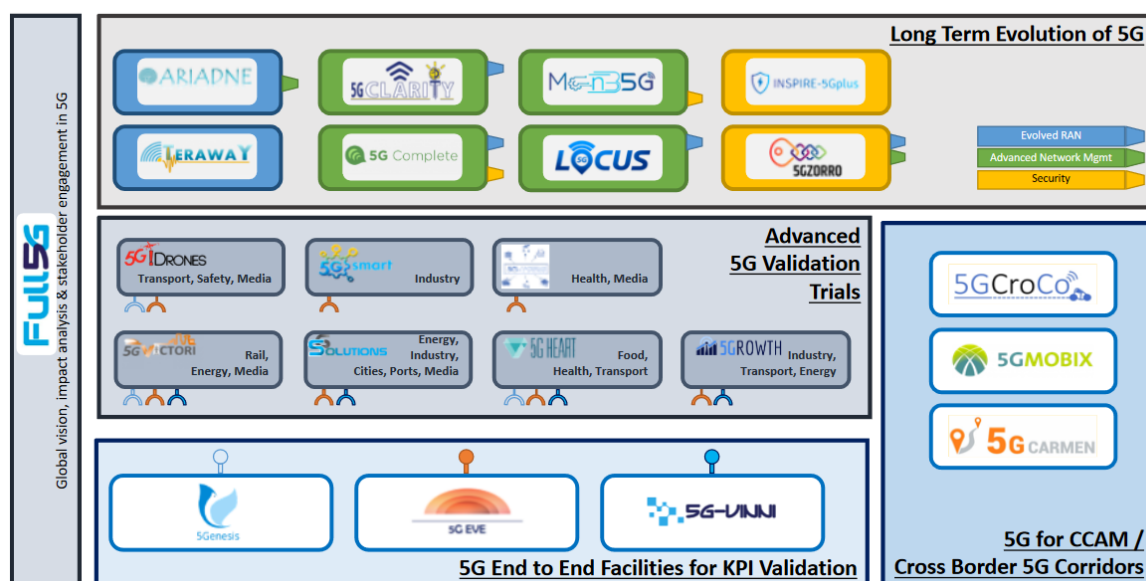


Figure 69. 5G-PPP Phase 3 Reference Figure¹²

5G-PPP Phase 2 and Phase 3 projects follow the overall Programme's goal to move from initial research results to large scale test-beds, getting closer to market applications. Since Phase 1, 62 projects in total have been so far contractually active in the 5G-PPP Programme, ensuring an outstanding momentum and dynamism. Also, note that Phase 2 Key Achievements from 5G-PPP projects include 60 highlighted results categorised under 14 program level achievements whereas a latest counting of Key Achievements v3.0 (Figure 70), including an updated list of key achievements from Phase 2 projects and key achievements from Phase 3 projects, amount to 80 innovations under 11 categories.

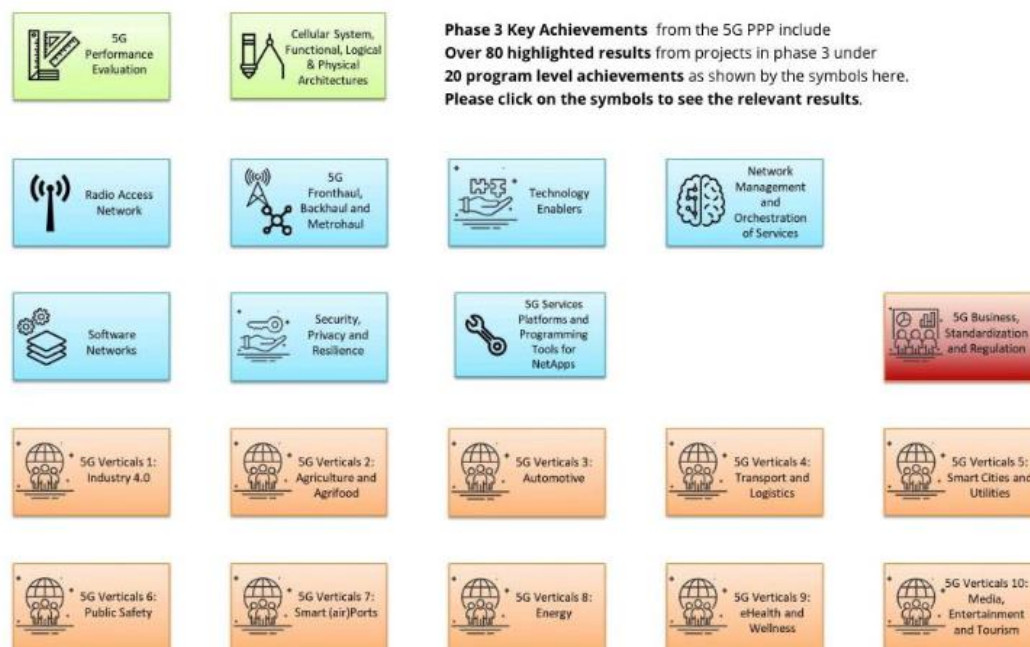


Figure 70. 5G-PPP Key Achievements v3.0

5G-PPP projects are detailed in: Annex 3 - 5G-PPP Progress Monitoring Report <https://bscw.5gppp.eu/pub/bscw.cgi/d424095/5G%20European%20Annual%20Journal%202021.pdf>

¹² 5G-PPP Progress Monitoring Report https://5g-ppp.eu/wp-content/uploads/2021/09/5G-PPP-PMR2020_Final.pdf

3.4. Review of current approaches in selected verticals

The use of edge-cloud technologies in specific sectors such as robotics, manufacturing, ports, machinery, telecom and renewable energy has been increasingly accepted due to the growing volumes of data generated over the last years.

An important number of companies belonging those sectors rely on traditional remote clouds in order to host operational data, but its increasing volumes are currently generating not only latency or transfer-speed related problems, but also higher costs, vulnerabilities, downtimes or loss of data, among others.

In order to tackle the challenges presented in the previous lines, an important effort have been recently done in edge-cloud technologies field applied in robotics, manufacturing sector, maritime ports, machinery construction, telecommunications and renewable energy. In the following sections, current approaches regarding edge-cloud technologies in aforementioned sectors will be discussed.

3.4.1. Edge-cloud technologies in robotics and manufacturing sector

Since its introduction in 2011, the so-called “Industry 4.0” [EMS-1][EMS-2] has widely exploited the concepts of edge and cloud technologies, even if not always properly named.

The punctual concept of “Cyber Physical System” (CPS), namely a device like a sensor or an actuator with intrinsic computational and network capabilities [EMS-3][EMS-4], introduces indeed the formal capability to perform computational operation at the edge of the so-called “automation pyramid” [EMS-5]. This allows to delegate to the lower layers of the network infrastructure simple (but eventually frequent) operations such as data filtering and structuring, saving computational and implementational time to the upper layers of the pyramid, and, at the same time, inputting them with cleaner data, increasing their performances.

For what concerns the cloud part,

3.4.1.1. The Automation Pyramid

A great relevance inside the Industry4.0 paradigm is covered by the so-called Automation Pyramid, based on IEC 62264 [EMS-5], which itself is based on the ANSI/ISA 95 [EMS-6] standard, evolution of the Purdue Enterprise Reference Architecture (PERA) model [EMS-7]. The pyramid structure, explained in Figure 71 expresses the typical hierarchy model of a production system before the Industry4.0 paradigm, when their relations were based upon the input received from the above level of the information system and the output given to the beneath one.

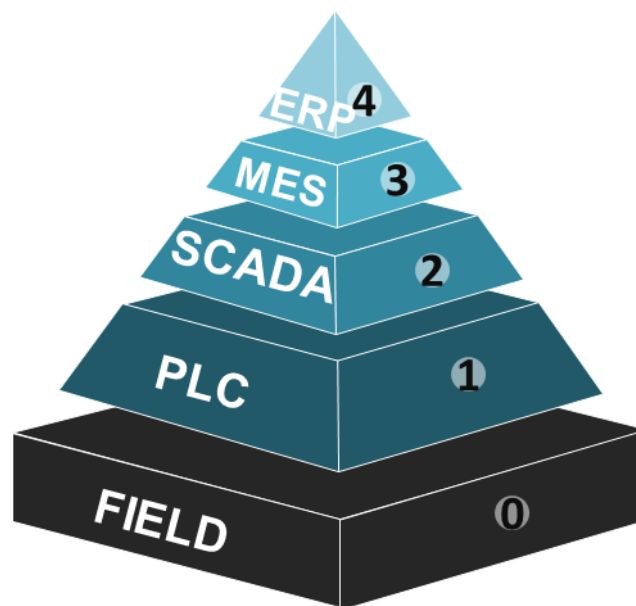


Figure 71. The automation pyramid

The four levels (from bottom to top) explicated in the figure are:

1. PLC/RTU: it represents the control which directly imposes signals to the assets' actuators and directly receives measured signals from the transducers. It's generally identified with Remote Terminal Units (RTUs) or Programmable Logic Controllers (PLCs), which close the stricto sensu control loops in manufacturing machines
2. SCADA: it represents the software demanded to supervision and synchronization over the under layer; it also has to gather and aggregate data to be provided to the upper level. These functions are commonly guaranteed by SCADA systems.
3. MES: it represents the application designed for the production control. It doesn't control the process, but it monitors production and targets by tracking the products, it schedules the resources and instructs the level beneath accordingly to the production targets. This level generally coincides with the Manufacturing Execution System (MES), whose operative functions are defined by the Manufacturing Enterprise Solutions Association between 2005 and 2013 [EMS-9].
4. ERP: it represents the application set devoted to the business and production management, which integrates modules and functions to schedule production, manage the supply chain, budget and manage projects. These functions are included in Enterprise Resource Planning suites, which communicate with the beneath software reading databases filled by the layer 3.

In addition to these layers, another one can be detected at the base of the pyramid: the so called "layer 0", which includes the hardware involved in the production process. Whenever this hardware has onboard electronics and logics enabling any addressing from the upper layers, the layer can be referred to the Cyber Physical System notation [EMS-4] the above-mentioned pyramid can be tilted to the one of Figure 71, according with a different perspective which allows every layer of the hierarchy to communicate with the layer 0.

This new representation highlights the connection that allows every software of the pyramid to gather directly the data needed to convey all the information required to quantify those Key Performance Indexes needed by the decision-making process.

3.4.1.2. The Cyber Physical Systems

According to Baheti and Gill [EMS-4] "the term Cyber Physical System describes a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities".

The need of a technology compliant with this definition takes the stage from an issue in the manufacturing world, where the control system was designed independently from the hardware/software to be controlled and then had to be ad hoc tuned through extensive simulations. However, this method has always been costly and time-consuming for complex systems made of subsystems acquired from different suppliers, because of the need to receive and compute signals gathered from devices by different manufacturers. For Original Equipment Manufacturers themselves, traditionally the strongest threat is to provide components able to easily integrate in their customers' products.

In this perspective, the biggest issue is represented by the fragmentation of research subjects, whose results are hard to integrate: typically, a formalism represents either the physical or the cyber part of a system, but not both. For example, a physical process is often modelled through differential equations, while a control flow can be represented through Petri nets or finite state automata. This separation implies a severe threat for verifying the correctness and safety of designs at the system level as well as the component-to-component physical and behavioural interactions [EMS-10].

As stated by Baheti and Gill [EMS-4], the main direction to follow in order to fulfil the research requirements is the one which develops innovative approaches to abstraction and architectures enabling seamless integration of control, communication and computation.

3.4.1.3. Above the Cyber Physical System

The first issue to solve to accomplish the aforementioned statement becomes placing the Cyber Physical System inside a structured architecture, completing the logic-layered pyramid through technological means able to build a defined communication between the layer 0 and the above ones.

End-to-end solutions are often to be discarded, since an enterprise usually runs hundreds or thousands of applications, which could be custom-built, acquired from a third party or parts of legacy systems (e.g. SAP): all these applications, to which websites and individual services developed for different departments have to be added, need to communicate with the Cyber Physical System, according with the "tilted" pyramid of Figure 72.

At the same time, approaches including the reduction of the applications' number have to be a priori discarded, since it's not feasible to design a unique software accomplishing all the features required to an enterprise and the applications' fragmentation gives IT managers flexibility to select the best solution to their particular purposes.

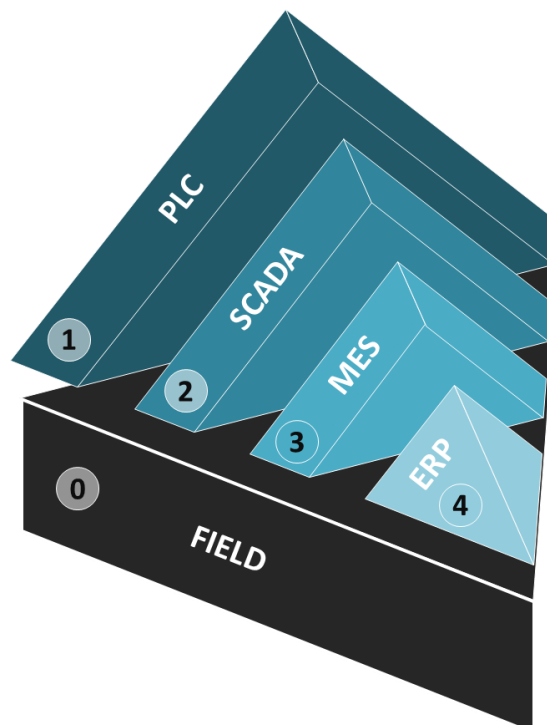


Figure 72. "Tilted" pyramid

Hence, during the last 15 years, the research and the market moved towards solutions able to gather and make data available among different applications, and hundreds of Enterprise Service Bus and integrated Platform as a Service were born and sold.

3.4.1.3.1. Enterprise Service Bus

According to Menge [EMS-11], the question about the definition of an Enterprise Service Bus (ESB) is hard to be answered, since it was coined by a Gartner analyst in 2002 to express the need for an infrastructure combining Message-Oriented Middlewares (MOMs), web services, transformation and routing intelligence as a backbone for Service-Oriented Architecture (SOA) [EMS-12].

This definition takes the birth from two different technologies:

- Service-Oriented Architecture (the idea): is an architecture concept which defines that application tools have to provide their business functionality in the form of reusable services. These services are generally self-contained and stateless business function accessible through a standardized, implementation-neutral interface. They are used by other applications which could also be implementations of services. With this approach complex processes are implemented through the so called "orchestration" of several services.

- **Message-Oriented Middleware (the mean):** substantially a message broker, it gathers messages so that senders and receivers can communicate without an end-to-end connection. The main threat concerning this approach consists in the fact that the middleware often uses proprietary protocols, leading to problems with Message-Oriented Middlewares of alternative vendors.

This leads Menge to define ESB as “an open standards, message-based, distributed integration infrastructure that provides routing, invocation and mediation services to facilitate the interactions of disparate distributed applications and services in a secure and reliable manner”.

Hence, to fulfil the requirements, the ESB has to provide some invocation features [EMS-13] (sending requests and getting responses) to receive and forward data (usually in XML format). This means the ESB has to support communication standards for web services (SOAP, WSDL, UDDI...) but has also to implement APIs for the communication with the Message-Oriented Middleware (JMS, JCA, JSR/JBI...) [EMS-14].

3.4.1.3.2. Integrated Platform as a Service

Even if the so called integrated Platforms as a Service (iPaaS) are fully compliant with the concept of Service-Oriented Architecture they have been having their own life since 2010, when a Gartner report named the concept for a change. They have been defined by Marian [EMS-15] as suites of “cloud services aimed at addressing a wide range of cloud, B2B and on-premises integration and governance scenarios, enabling development, execution and governance of integration flows connecting any combination of on-premises and cloud-based processes, services, applications and data within individual, or across multiple, organizations”.

The characteristic defining these type of platforms apart from other SOAs ones is essentially their running on the cloud. This feature enabled a selling strategy allowing B2B customers pay-per-computation or pay-per-storage policies, spreading the SOA paradigm also to businesses characterized by high peaks of data in limited time periods, for whom on premise services are not cost-effective.

For these companies, platforms both from legacy market (Dell Boomi, SAP Hana, Jitterbit), both on the open source side (softMule) are significantly increasing their business selling cloud services for protocol bridging, messaging transports, transformation, routing, service virtualization, adapters, orchestration, registry, repository, partner community management, MFT, development tools and others (remark also the 34-billion acquisition of Red Hat by IBM).

3.4.1.4. System Integrators

The technologies mentioned above require a high effort to translate/adapt/reroute protocols in use among the different layers of the automation pyramid of Figure 72. In particular (for quantitative reasons), the effort is focussed on interfacing all the elements of the Cyber-Physical System with the middleware and on designing data models for the messaging/storing of production and logistic information.

A huge threat for System Integrators is hence the modular architecture of the facilities, with different architectures and data models for different departments: this issue affects not only the level 4 of the automation pyramid of Figure 72 (which has to interface with different Manufacturing Execution Systems for different production departments) but also the internal logistic system, which has to be warned about readiness of input and output for every module of the production chain.

3.4.2. Edge-cloud technologies in maritime port sector

3.4.2.1. Introduction and Motivation

For a port terminal to become fully automatic, machinery must work without a driver in the cabin (although human-in-the-loop supervision or remote control is expected). However, the automation of the physical handling (unloading, storing, loading) of containers has only been partially achieved. After more than 25 years of developments, robotization has definitively taken off and more than 1100 driverless cranes are in operation worldwide and thousands of Automated Guided Vehicles (AGV's) carry out transport operations from quay to yard, becoming a standard product in modern terminals, but all these automated robots are only placed in 35 out of approximately 2000 container terminals globally (1.75%). The automation of quay (ship-to-shore) crane is

less developed, as current practice requires that controlling their dynamic behaviour, such as undesirable swaying, is the responsibility of a skilled operator [EMP-1–5].

Some of the main limitations for this successful deployment of full automated CHE comes due to the requirements of deploying a high variety of sensing systems (inertial sensors, ultrasonic sensors, eddy current sensors, radar, lidar, imaging sensors, buried in the ground or with antennas in the bottom of the vehicle) in order to support tasks such as container positioning, detection, and handling using computer vision methods or corner casting recognition [PA-6]. Connecting all these sensors over the internet is a challenge as container terminal environment are inherently hostile for wireless communication. Furthermore, to support remote controlling operations from a control room, cranes should be equipped with multiple high-definition cameras (can vary from 6 to 27 cameras, depending on their size and payload capabilities), leading to a total uplink bandwidth of approximately 30 – 120 Mbps. Large coverage requirements are also imposed for enabling cranes movement within terminal ports (e.g., RTG cranes have up to 1 km range of mobility with speeds up to 40km/h).

Although wireless technologies have been widely used for many years in container terminals for non-time-critical communication, the connectivity challenges for automation or remote controlling initiatives have been fulfilled to an extent by a mix of fixed and wireless networks, using fibre-optic cables together with Wi-Fi and 4G systems. However, on the one hand, fibre solutions require expensive and time-consuming deployments, as well as some areas of ports are unreachable via wired solutions. On the other hand, wireless Wi-Fi and 4G technologies are not sufficient to cope with ultra-reliable and low-latency communications requirements of automation (e.g., Wi-Fi only delivers a coverage area of tens of meters with limited QoS or switching between multiple APs can take several seconds). 5G, unlike 4G, is expected to provide significantly higher bandwidths, both in the downlink, and more importantly in the uplink, and a rapid response rate to the controller. However, even though, 5G networks on their own will not guarantee such ultra-low latencies, as all mobile data is sent to the operators' core network before reaching an external data network, significantly adding the overall latency.

The advent of edge computing deployed at local gateways will have a twofold advantage:

1. Through user plane and control plane separation, edge computing ensures the data is kept being processed locally within the port networks, thereby reducing the overall latency.
2. Edge computing can create a private local network, improving data security. Given that ports are independent enterprises, the port authorities will not want their data to interact with the MNOs external infrastructure.

3.4.2.2. Edge computing technologies for maritime port sector

Regarding edge computing solutions, Dell and Intel are leading the market race, helping to different stakeholders across the globe to develop, test, and deploy the edge computing technologies to make the vision of maritime automation a reality, enabling maritime organizations to build Edge to Cloud infrastructure that adapts and scales to help port operators to sustain, grow, and protect their data, cargo, workers, environment and ultimately their business. A brief portfolio of Intel-Dell solutions is depicted below.



Figure 73. Dell-Intel edge and IoT portfolio for port operations [EMP-6].

3.4.2.3. Edge computing uses cases for the port

Next, different edge computing applications for port automation are briefly described.

- Fleet and asset management solutions could use Edge computing hardware and software to increase the visibility, integrity, and security of assets moving through ports' premises, helping operators to gain near real-time tracking and monitoring of asset location, temperature, humidity, tire pressure, oil and fuel status, and maintenance in general. Comprehensive dashboards enable effortless monitoring and analysis; and may include programmable notifications and alerts for quick intervention.
- Machine vision systems used for container identification are expensive due to the dedicated monolithic architecture (tightly coupled HD cameras and image processing servers in the far cloud). A more cost-effective machine vision system would offload the image processing capability from cloud servers to local edge computing servers.
- As another example edge computing use can benefit port's networks, making future upgrades and daily maintenance easier, as well as facilitating the AI and big data algorithm training by breaking the data silos [EMP-7].

A more detailed description of ML applications on the port, and how they can be used by means of edge computing solutions is described next.

- **Quayside ML:** The performance of quayside planning depends on many factors, including vessel arrival times, vessel call patterns, peak demands, and the handling capabilities of the quayside equipment. Uncertainties may result from a lack of reliable information and forecasting. To limit some of these uncertainties, strong research has been focused on the analysis of satellite Automatic Identification System (S-AIS) data. It will help for identifying patterns and anomalies of vessel operations, e.g., to avoid vessel accidents or to identify unauthorized activities like illegal bunkering. Applications of ML in the quayside include *Prediction of vessel arrival times*, *Prediction of turnaround times*, *Prediction of ETC time*, *Berth planning*.
- **Yard ML:** Several complex planning and optimisation problems result from yard operations (e.g., yard allocation, post-stacking, crane scheduling, etc.) It is important therefore to reduce uncertainties by predicting future scenarios by making use of ML applications like:
 - **Prediction container dwell times:** Different algorithms have been developed and evaluated. Models can be used to assess the impact of changing determinants on the container dwell times yard capacity and terminal demurrage revenues.
 - **Container stacking:** algorithms have been developed to predict the quantity of incoming containers and weight groups of containers to optimise the container stacking policy.
 - **Predictive maintenance (PdM) systems:** can be applied to all types of yard cranes. It allows to predict the need for maintenance of these assets, anticipating failures and improving decision making. This results in the decreasing of machine downtime, costs, control, and an increase in quality of production. An in-depth systematic review of predictive maintenance has been carried out by [EMP-8].
 - **Computer Vision techniques:** can be used towards several objectives: ISO-code recognition [EMP-9], assist in the container-spreader alignment [EMP-10], adaptive container landing system [EMP-11].

It should be noticed that the two latter use cases will be implemented along aerOS project. While a regular ML model will be tested at first, the scope of both pilot studies is to pursue and obtain frugal AI models that can be deployed across the edge – cloud continuum of the project.

- **Landside ML:** Improving landside operations by ML can lead to better hinterland accessibility and inland connectivity, which is crucial for the competitiveness of container terminals. Contextual data extracted from already deployed sensors can be used to better understand and coordinate traffic flows, including *prediction of truck traffic*, *prediction of truck waiting and turnaround times*, or *prediction of truck delays*.

3.4.2.4. Relevant research initiatives

Many research projects in the port industry indicated a growing interest in automation technologies for the maritime industry. Some of them are briefly introduced next.

- **iTerminals4.0** [EMP-12] is one of many research projects co-funded by the EC. Its goal is to boost digitalisation of port operations, and the adoption of Industry 4.0 technologies within the container-handling, by means of an upgrade of port equipment's sensor networks, the design of advanced big data and predictive analytics, the application of AI, as well as the provision of business intelligence models and real-time dynamic KPIs reporting.
- The **COREALIS** [EMP-13] project proposes a strategic, innovative framework, supported by disruptive technologies and emerging 5G networks, for cargo ports to handle future capacity, traffic, efficiency, and environmental challenges.
- The **CYBER-MAR** [EMP-14] project aims to develop cyber preparedness for cyberattacks in the maritime environment and to estimate the impact of a cyberattack from a financial perspective.

Beyond European R&D projects, several private partnerships have been carried out in the latest years for speeding up port automation:

- The use of autonomous surface vessels navigating without human control forms part of project developed by Mitsui OSK Lines testing Rolls-Royce's intelligent awareness system in its vessels. The system combines data from onboard sensors with information from bridge systems looking for a safer, simpler, and more efficient way to operate [EMP-15].
- The port of Hamburg has created a Decision Support System (DSS) using deep learning techniques and neural networks capable of predicting the behaviour of land transport. The system forecasts the times when lorries should reach terminals and the drivers have received a notice about the expected terminal entrance times. The model supplies a dynamic forecast of the workload considering changes in the surrounding conditions like road and access route saturation, real ship arrival time, or degree of terminal saturation.
- The port of Qingdao in China and Ericsson launched a partnership programme at MWC 2019, following a technical trial in late 2018, to develop a 5G smart port solution. One of the key goals was to demonstrate the advantages and labour cost savings that could be possible if 5G networks were used for automation compared to a traditional port with no automation.
- The port authority of Livorno, together with Telecom Italia (TIM) and Ericsson has defined an innovative model to assess the introduction of 5G technologies and explore how digital transformation can meet the UN SDG-2030 goals [EMP-16].
- The engagement of Huawei with the port authority at Ningbo, one of the world's largest with over 550 gantry cranes, successfully demonstrated the use of 5G together with Edge computing, delivering high data throughput needed to serve many HD camera feeds, together with latency of less than 20 ms for vehicle remote control [EMP-17].

3.4.3. Edge-cloud technologies in mobile machinery sector

In this project TTC and John Deere are targeting to develop a High-Performance Computing Platform for Connected and Cooperative Mobile Machinery. This platform has the potential to reduce the CO2 footprint in areas like agriculture, construction, or forestry. The main motivation for this vertical is as follows. The digital transformation in agriculture, construction, where mobile machinery is used, has made significant progress over the last decade. Especially Precision Farming Technologies offer a pathway towards reducing inputs, maximizing yields and quality of produced goods. Digitalisation allows for integrated control of machines and vehicles involved in production processes. At the same time farming needs to interact with other production systems and information service in the food production and food value chain. The required network connectivity everywhere and always is still a challenge. Cellular networks may need a long phase of invest and deployment until a full coverage in rural areas is achieved. *Edge computing* in connection with locally limited and temporary networks will be needed as enabler for autonomous machine fleets.

Connected and cooperative agricultural mobile machinery is a key to synchronize and optimize the tractor work for productive and sustainable farming in the future. Due to the challenges mentioned above the existing systems are pushed to their limits, e.g. to perform data access and processing, ensure data privacy and security but using *the data also from cloud*, the control systems, in particular in-vehicle computing and networking platforms shall be modified and extended with the new components and modules. The proposed robust and flexible solutions need to provide a connectivity from machine to machine from everywhere in real-time for large-scale agricultural production system on one side, but also deliver certain real-time performance still navigating the overall system remotely and controlling (i.e. supervising) execution of the agricultural work process. The similar technical tasks are relevant for e.g. road building machinery. The application of *Cloud computing* might be interesting due to “convenient, on-demand network access to a shared pool of configurable computing resources”, see NIST definition of Cloud Computing [EMM-1]:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST proposes three service models in this regard: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) and four deployment models of cloud: private, community, public and hybrid. In the project we will perform an analysis which service model and deployment model should be considered in our use case. For instance, using cloud services for less critical services. TTConnect Cloud Service and IoT solutions by TTC (not a part of the aerOS project development but included in the TTC’s product portfolio) offer a connectivity technology enabling manufacturers of mobile machinery to monitor and manage all their vehicles around the world at any time with any web-enabled device – with only one off-the-shelf solution. The TTC’s offering includes a hardware unit named TTConnect Wave (IoT gateway), a M2M SIM card, as well as a cloud service platform and a web portal. TTControl’s IoT connectivity solutions are applicable for various use cases in the off-highway sector. Whether you need to analyze sensor data from your harvesters, protect your wheel loader from theft on a construction site or optimize the routes of your garbage truck fleet, one can benefit from TTConnect Cloud Service. Two application examples as below:

Agriculture: through access to vehicle data, manufacturers of agricultural vehicles – such as harvesters, balers or sprayers – can troubleshoot any failure in a short time. By analyzing the data collected by sensors via the in-vehicle CAN interfaces, operator needs can be anticipated and considered for the development of upcoming vehicle architectures.



Construction: by regulating the hydraulic systems, TTConnect Cloud Service helps you to avoid over-usage and misuse of your excavators, wheel loaders or rollers and lowers the mechanical stress of valves, pumps or motors. It increases the productivity of your machinery and allows for predictive maintenance. The key features of this solution are e.g.: over-the-air updates of complete machine software, fleet management and maintenance, collecting and analyzing machine data in real-time, simple and intuitive configurable web portal, creation and configuration of alarms, unparalleled machine integration with TTControl controllers and displays.

So far about Cloud. Using IoT services also edge computing is foreseen for more critical tasks like e.g. field borders in the overall system to e.g. to enable real-time control. For this automatic control as part of the vehicle system (automatic driving) is critical.

In the aerOS project TTC will focus on new electronic vehicle architectures and High Performance Computing Platform (HPCP) prototype to provide a connectivity from vehicle to vehicle from everywhere in real-time for large-scale mobile machinery system on one side, but also deliver certain real-time performance still navigating the overall system remotely and controlling (i.e. supervising) execution of the e.g. agricultural or construction machine work process. Like automated driving levels of autonomy of ADAS, the idea here is to develop a proof-of-concept solution for Partial or Full Automation performed by machine instead of a human.

According to the analysis of the state-of-the-art, there is still a need for further research to create high performance computing platforms, being able to host applications targeting SAE levels 3+. There is a lot of research (also in automotive) done on module and component level in hardware and software, which might be used in an integrated system. What is missing and covered within this project is the development of a computing platform, fulfilling requirements to be able to host automated driving functions on one hand and considering safety and dependability attributes on the other hand by applying already investigated patterns e.g., on system architecture level. For instance, the following applications/services can be deployed and executed on such as system / platform:

- Level 3 Highly Automated – Environment monitoring, AI and deep learning, Convoy (1 driver)
- Level 4 Offroad Autonomous (High automation and Most conditions) – Offroad Autonomous, Onroad autonomous or driver, Remote monitoring
- Level 5 Autonomous (Full automation and All conditions) – No Driver, Onroad + Offroad Autonomous, i.e. “Hands Off”, “Driver Off”.

This use case proposed in aerOS will contribute to enabling sustainable mobile machinery solutions for energy optimisation and noise reduction. The data from sensors (e.g. cameras, LIDAR, Radar) as well as operating instructions from a cloud will be safely and securely processed to feed a grid-connected electric swarm. *Cloud to cloud interoperability* will be adopted for the optimization of the data used to remotely control the swarm of vehicles. The developed solution will be capable to e.g. perform computational tasks in support of demonstrating fully electric swarm of vehicles safely and securely operating e.g. in platooning or other swarm combinations. The solution will bring higher performance and connectivity capabilities vs. existing solutions brought to the mobile machinery. Using IoT services also edge computing is foreseen for more critical tasks like e.g. field borders in the overall system to e.g. to enable real-time control.

3.4.4. Edge-cloud technologies in telecom operators sector (a usability perspective)

The digital transformation trends across most industries exhibit growing adoption of enabling technologies such as cloud, edge, AI and IoT. In this landscape, next generation networks that offer reliable data transport, compute at the edge, and automation for mass connected assets and devices, become the backbone of new use cases, such as industrial asset monitoring and digital twins enabled by sensors. According to the recently published (September 2022) GSMA Intelligence report for the IoT and Enterprise [ETS-1], network operators see the enterprise use cases as the incremental opportunity to increase revenues outside the very-competitive-low-profit-margin telecommunications market and are expanding their connectivity services portfolio with other digital services such as Cloud, IT, IoT, security among other professional services.[ETS-2] Indicatively, based on the analysis of eleven major operators, the report reveals that the average contribution of enterprises services to total telco revenues has reached 30% in 2020, and there is still significant room for growth. Furthermore, seeking to monetise their investments in 5G, MNOs (Mobile Network Operators) promote the edge computing and massive IoT as the 5G value proposition towards their enterprise customers. A clear advantage comes from the fact that MNOs’ points of presence are unique in addressing the proximity requirements of most demanding use cases with deployment options ranging from deep and far edge (up to 5km and 10km from end user respectively) to aggregated edge (up to 30km) [ETS-2].

In the digital transformation directives, private networks are gaining momentum responding to the modern network’s industries mandates, and are quickly becoming a multi-stakeholder game, raising the urgency for the operators to prepare and act fast. The GSMA enterprise survey [ETS-1] across most vertical industries on “who would you prefer to partner with to create a private network”, shows that network operators are the first partner of choice only for the 24% of the responders while the majority (50%) declare preference towards infrastructure/hardware vendors. At the same time, all three major hyperscalers, Amazon (AWS), Microsoft (Azure) and Alphabet (Google Cloud), taking advantage of their cloud computing capabilities, have expanded their portfolios with their own flavours of private 5G, and most have completed strategic acquisitions and hired from the mobile industry [ETS-1]. Against this treat, operators are preparing to respond with slicing and edge computing, capitalising on the **5G SA (5G Stand Alone) network architecture** that inherently supports the digital transformation needs. 5G SA, with its Service-based Architecture and cloud-native functions, and the

advanced functionalities such as network slicing and Massive Machine type Communication (MMTC), Multi-Access Edge Computing (MEC), is a key enabler for the enterprise edge and IoT solutions.

At the same time, the exponential increase of the number of connected devices and volume of data handled by the network have significantly increased the energy consumption of telecommunications networks that is becoming an extremely critical factor [ETS-3]. According to GSMA [ETS-3], energy consumption is one of the highest operating costs for network operators typically covering 30% of operations expenses (OPEX). At the same time, the 5G networks are expected to account for 21% of the total energy consumption by 2025. Turning off equipment when not in use, even for a short time, and putting some network resources in standby mode, reducing the site infrastructure are important energy saving actions. It is anticipated that through the use of **AI/ML mechanisms**, network behaviour can be predicted and controlled intelligently, leading to unified, automated management of resources and efficient networks' reconfiguration, that can quickly adapt to changes on demand and reduce the energy consumption by ensuring the accurate use of resources as necessary to guarantee the performance levels requested per case.

Concerning the evolution path, in the past years, most operators have implemented a strategic agenda towards transition from physical network infrastructure to cloud-based architectures, investing in NFV network based on cloud technology such as OpenStack¹³[ETS-4], VMWare¹⁴[ETS-5] among other best practices [ETS-6]. Presently, European operators are progressing with the 5G rollouts and network modernisation. There are nearly 200 live 5G networks in seventy countries, including 68 operators providing 5G Fixed Wireless Access (FWA) services and 23 delivering Stand Alone (SA) 5G services. According to the GSMA, 5G connections will surpass 1 billion in 2022 and by the end of 2025, 5G will account for over a fifth of total mobile connections, and more than two in five people globally will live within reach of a 5G network [ETS-7]. In parallel, interest is raising on the deployment of stand-alone (SA) 5G networks that are expected to pave the way for edge-cloud adoption. It is noteworthy that 5G SA services in Europe are now available in Finland, Germany and Italy and more deployments are expected in the next few years [ETS-7]. On the sustainability front, European operators are at the forefront of adopting cutting-edge, energy-efficient technologies and the use of renewables, with many already reaching 100% renewable electricity use across their footprints, powering their network infrastructure, data centres and other sites [ETS-8].

In conclusion, from the telecom operator's usability perspective, it becomes evident that the use of edge-cloud technologies is pivotal in all dimensions:

- As a technology supplier, to assume the role of edge-cloud provider and offer enterprise, beyond connectivity, services, supporting the vertical industries' digital transformation and capitalising the 5G network investments
- As a technology consumer, in the course of digitalisation and operating expenses reduction, to exploit technology towards its own transformation, at the business level and for the network sustainability. In this perspective, use cases such as smart, energy efficient buildings become attractive to be deployed in own telecom premises.

The technological ecosystem and the involvement of key players and Standards Development Organisations (SDOs) towards the edge-cloud implementations in the mobile networks domain are depicted in *Figure 74*, and are detailed in the subsections that follow. In a highlight, two standardisation groups (3GPP and ETSI) and two industry fora (GSMA and TMFORUM) are taking the lead in building of the edge-cloud in telecommunication business.

¹³ <https://www.openstack.org/use-cases/telecoms-and-nfv/>

¹⁴ <https://telco.vmware.com/>

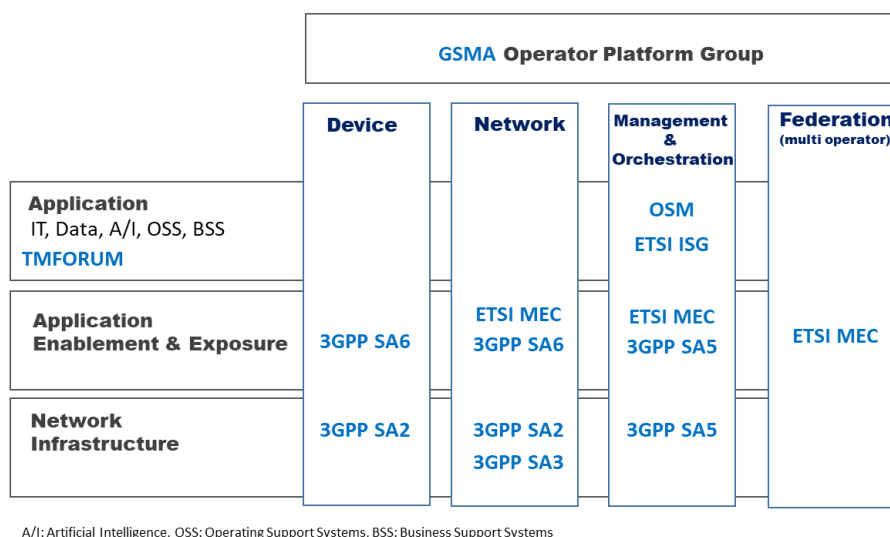


Figure 74: Overview of the Involvement of SDOs for Edge Computing in Mobile Networks, inspired by [ETS-9]

3.4.4.1. 3GPP for Telecommunications Edge Cloud

The 3rd Generation Partnership Project (3GPP)¹⁵[ETS-10] unites seven telecommunications standard development organizations, providing their members with a stable environment to produce the Reports and Specifications that define the 3GPP system, covering cellular telecommunications technologies, including radio access, core network and service capabilities. 3GPP with the most recent 5G standards aims to develop features that go beyond typical end-consumer expectations (e.g. higher speeds, better coverage), and towards capabilities that enhance the communications for vertical industries such as public safety, automotive, drones, factories of the future, IoT, in sync with the advent of the industry 4.0 revolution [ETS-11]. The 5G SA Architecture is a key enabler for the edge-cloud momentum, introducing fundamental concepts such as SBA (service-based architecture) that empowers virtualisation and intelligent distribution of network functions at the edge, and slicing, enabling on-demand, user-driven and of guaranteed quality services.

The 3GPP architecture working-group (SA2) has specified the overall 5G system architecture, detailing features, functionality and services and the 5G SA capabilities were gradually introduced in the specifications' Release 15 (frozen in 2019), Release 16 (frozen in 2020) and Release 17, Release 18 (up to the time of writing open). Highlight 3GPP developments that unleash the capabilities to support the intelligent edge-cloud era include:

- **Network Slicing**, key feature of 3GPP TS23.501 [ETS-12] 5G System Architecture, is a concept for running multiple logical and customised networks on shared common infrastructure, with agreed SLAs and requested functionalities. There are many parallel initiatives in the definition of the end-to-end slicing with fundamental concepts being the resource model, service profile and management in 3GPP TS28.541 [ETS-34] and GSMA's Generic network Slice Template (GST)[ETS-13].
- **Network Exposure Function (NEF)** 3GPP TS 29.522 [ETS-14], that introduces the "Network Programmability" concept allowing the development of network-aware applications that can adapt to network conditions and interact requesting dynamic network reconfigurations and quality of service adaptations.
- **Network Data Analytics Function (NWDAF)** 3GPP TS 29.520 [ETS-15] that exposes insights to the core network data by streamlining the way they are produced and consumed to enhance end-user experience.

Edge computing in particular has been a major focus in 3GPP Release 17, with four key groups in TSG SA (Technical Specification Group System Aspects) carrying out related studies and normative work [ETS-16] as follows and graphically depicted in *Figure 75*:

¹⁵ <https://www.3gpp.org/>

- SA2: System Architecture enhancement for supporting Edge Computing.
- SA3: Security aspects for supporting SA2 and SA6 architectures.
- SA5: Management & Charging aspects on Edge Computing.
- SA6: Edge Enabler Layer architecture, and deployment scenarios

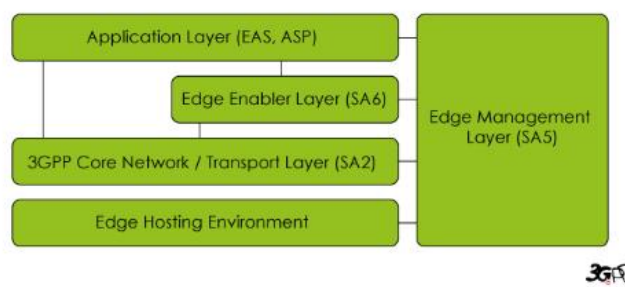


Figure 75: Simplified View of the 3GPP WGs Edge Work

Fundamental to facilitating edge-cloud deployments is the work led by SA6 on application-enabling service frameworks (also responsible for vertical enablers and mission critical services), that has delivered the following specifications:

- **Common API Framework for 3GPP Northbound APIs (CAPIF)** 3GPP TS 23.222 [ETS-17] standardises common capabilities exposed by 5G SA Northbound APIs, for a variety of processes, such as on-boarding/off-boarding Application Functions, service discovery and management, event subscription and notification, security and charging.
- **Service Enabler Architecture Layer for Verticals (SEAL)** 3GPP TS 23.434 [ETS-18] specifies a functional architecture to support vertical applications by specifying common application plane and signalling plane entities and services (namely group management, configuration management, location management, identity management, key management and network resource management) that can be shared across vertical applications.
- **Architecture for Enabling Edge Applications (EDGEAPP)** 3GPP TS 23.558 [ETS-19] builds upon the concepts set by CAPIF and SEAL and describes the enabling layer and application architecture to implement edge applications on the Edge Data Network (EDN). The enabling layer refers to the exposure of the northbound APIs towards the edge applications, integration with the 3GPP Network and the communication of the application clients running on the UE with the Edge Application Servers (EAS) deployed on the EDN, including capabilities such as service provisioning, rich application discovery and service continuity. Additionally, 3GPP TS 28.538 [ETS-20] focuses on Edge Computing Management, addressing Lifecycle management (e.g. on boarding) of Edge applications. GPP TS28.552 [ETS-21] and TS 28.554 [ETS-22] define the performance measurements and KPIs for edge applications respectively.

3.4.4.2. ETSI MEC for Telecommunications Edge Cloud

The Multi-access Edge Computing (MEC) initiative is an Industry Specification Group (ISG) within ETSI [ETS-23]. The work of the MEC initiative aims to unite the telco and IT-cloud worlds, providing IT and cloud-computing capabilities by specifying the elements that are required to enable applications to be hosted in a multi-vendor multi-access edge-computing environment. MEC also enables applications and services to be hosted ‘on top’ of the mobile network elements, and benefit from being in close proximity to the customer and receiving local radio-network contextual information. ETSI ISG MEC specified a common and extensible application enablement framework for delivering services, specific service-related APIs for information exposure and programmability, as well as, management, orchestration and mobility related APIs. These APIs facilitate the running of applications at the correct location at the right time and ensure service continuity.

ETSI ISG MEC is currently studying MEC federations to enable shared usage of MEC services and applications across MEC systems in support of a multi-operator/multi-network/multi-vendor environment [ETS-9]. It is noteworthy that in the initial scope of the initiative, MEC stand for Mobile Edge Computing based on 3GPP access-related technologies, and in a second phase, it extended to Multi-Access Edge Computing including Wi-Fi and fixed access technologies.

ETSI ISG has defined a number of specifications [ETS-24], among which we can highlight the following:

- Multi-access Edge Computing (MEC); Framework and Reference Architecture MEC 003 [ETS-24]

- General principles, patterns and common aspects of MEC Service APIs MEC009 [ETS-25]
- Study on Inter-MEC systems and MEC-Cloud systems coordination MEC035 [ETS-35]

While MEC is a design characteristic of the 3GPP 5G Architecture [ETS-36][ETS-37], ETSI ISG MEC alignment with 3GPP SA2 & SA6 is on-going [ETS-38] including aspects related to MEC 5G Integration and future evolution, MEC Federation as well as addressing the operator requirements as set by the GSMA OPG (Operator Platform Group). A very important result of this synergy is the publication of a Harmonised Edge Computing Architecture, as presented in *Figure 76*, to be used as a blueprint for edge-cloud deployments for the telecom business.

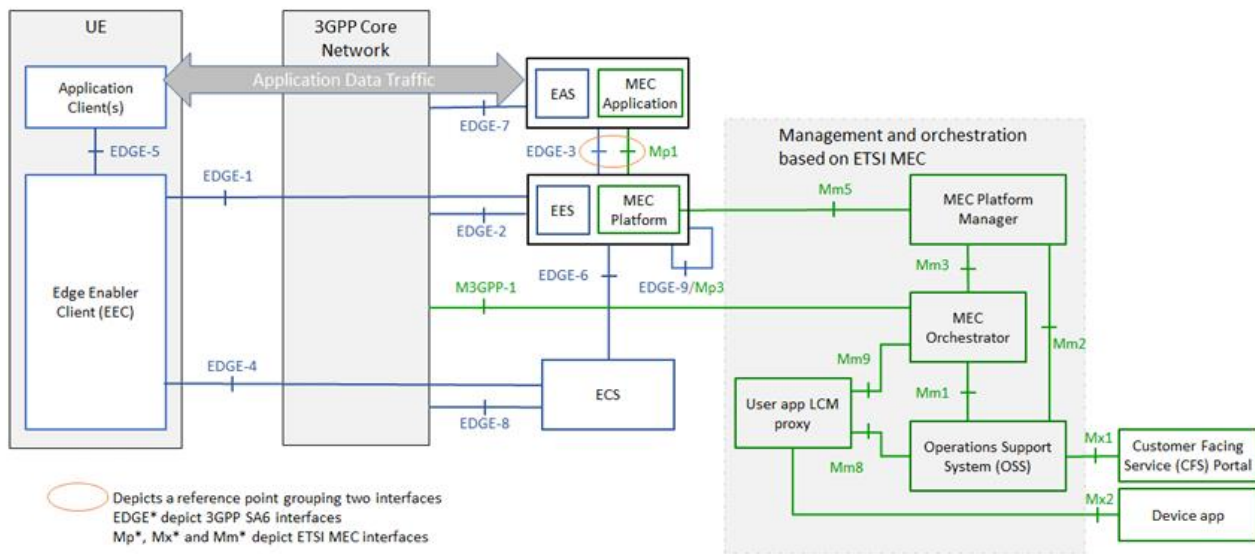


Figure 76: Synergised Mobile Edge Cloud Architecture Supported by 3GPP and ETSI ISG MEC specifications [ETS-9]

3.4.4.3. GSMA for Telecommunications Edge Cloud

GSMA [ETS-36] is a global-led organisation within the mobile industry, including 750 mobile operators and over 400 companies in the broader mobile ecosystem, and aims to drive initiatives shaping the future of mobile communications with invaluable insights and industry intelligence in the priority topics of 5G, IoT, Fraud and Security.

Within its Future Networks thread, there are two highlight initiatives related to the edge-cloud momentum for the telecommunications industry, the Operator Platform Group (OPG) and TEC (Telco Edge Cloud) Forum:

- **GSMA OPG & OPAG:** The **Operator Platform Group (OPG)** [ETS-26] seeks to support operators to monetise their vast local footprint, their existing relationships with enterprises, and their competence to provide high-reliability services supporting the digital sovereignty through 5G, by bringing in the missing piece, which is the ability to package and expose their networks in a scalable fashion across multiple operators. OPG works on defining a common platform that exposes operator services/capabilities to customers/developers in the 5G-era in a connect-once-connect-to-many model. OPG is open to the wider edge-ecosystem and brings together operators, platform developers, edge cloud providers, Standards Developing Organisations (SDOs), Open-Source Projects, industry 4.0 and market participants. It targets to create the architecture and technical requirements to guide other Standard Developing Organisations (SDOs) in the development of specifications, in the first phase with focus on the Edge, and in future expanding with other capabilities such as slicing. While the Operator Platform Group (OPG) is responsible for the technical requirements and the development of the operator platform, a special subgroup, the **Operator Platform API Group (OPAG)** [ETS-27] undertakes the alignment in Operator Platform (OP) APIs fulfilling OP requirements and the collaboration with and contributions to SDOs (such as 3GPP, ETSI and Linux Foundation).

The OPG project considers the enhancement of the Edge capabilities with [ETS-27]:

- Smart Edge allocation, and selection to perform load deployment and access from the closest edge
 - Edge federation to offer a multi domain Access to customer and enhance edge service under roaming scenarios
 - Tight network integration to enhance mobility and user experience
- **TEC Forum:** The Telco Edge Cloud (TEC) [ETS-28] Group has a commercial focus and brings together over 20 operators, covering all regions, who are working to promote a collaborative deployment of cloud capabilities at the edge of their networks. TEC is aiming to align Multi Access Edge Computing (MEC business models, charging principles and commercial deployment considerations), and has primary focus on edge cloud trials and POCs [ETS-29].

3.4.4.4. TMForum for Telecommunications Edge Cloud

TMForum is an widely accepted alliance of 850+ global companies working together to break down technology and cultural barriers between digital service providers, technology suppliers, consultancies and systems integrators [ETS-30]. The focus of the work is to help Communication Service Providers (CSPs) towards their digital transformation journey through various ways, from managing the process of transformation through a maturity model, to offering practical toolkits, widely-adopted frameworks including Open APIs, as well as, accelerating innovation through rapid POC cases. As of 2020, strategic collaboration of TMForum with GSMA was announced¹⁶[ETS-31], building upon the realisation that GSMA's focus on mobile networks and TM Forum's efforts in IT, data and AI make the collaboration between the pair an obvious fit.

TMForum is putting effort in both edge computing and Network-as-a-Service (NaaS) and the most relevant work is highlighted below [ETS-32]:

- **Open Digital Architecture** [ETS-33]: An important initiative towards replacing traditional operations and business support systems (OSS/BSS) with a new approach to building software for the telecoms industry, opening a market for standardized, cloud-native software components, and enabling communication service providers and suppliers to invest in IT for new and differentiated services.
- **Open APIs** [ETS-39]: Core to the development of the cloud-native Open Digital Architecture are the OpenAPIs, a suite of application programming interfaces that enable services to be managed end-to-end throughout their lifecycle within an environment where multiple partners might be involved. As an example, API Suite Specification for NaaS [ETS-40] supports exposing and managing "Network" Services while more than sixty (60) REST-based Open APIs are developed collaboratively. Up to date, 141 of the world's leading communications service providers (CSPs) and technology ecosystem participants have signed the Open API Manifesto publicly demonstrating their endorsement of TM Forum's suite of Open APIs.
- **Becoming EDGY** Catalyst Project [ETS-32]: The awarded 2019 best new Catalyst in show project sets to explore the maturity of the edge-cloud management solutions in the market to build the ability of dynamic network slicing with zero-touch orchestration, a critical success factor for 5G.
- **The 'EDGE' in Automation** Catalyst Project [ETS-32]: Building upon the findings of EDGY, this project demonstrates solutions for Edge Compute as a Service (ECaaS). It aspires to deliver to event developers/suppliers (e.g. concert, sporting/gaming event, exhibition, etc.) pre-scheduled ECaaS packages (e.g. capacity, image recognition, surveillance, doors lock/unlock, emergency services, etc.) at a venue when a crisis occurs, requiring real-time reconfiguration of the edge to deliver public safety emergency services.

TMForum, from the standpoint of aiding the telecoms transformation process, is explicitly monitoring, promoting and sharing the sustainability initiatives of the CSPs as a separate topic [ETS-41]. As a highlight, Catalyst projects of interest include:

¹⁶ <https://www.mobileworldlive.com/featured-content/home-banner/gsma-chair-opens-door-to-tm-forum-collaboration/>

- **5G greener telco Catalyst Project:** The 2020 Catalyst award winner for impact for society, builds upon the fact that the necessary doubling of the number of (5G) base stations and the introduction of large-scale Multiple Input Multiple Output (MIMO) technology have led to a significant increase in 5G power consumption, 2.5 to 3.5 times higher than for 4G base stations. The project proposes the development of energy saving solutions including coordination between 4G and 5G, 5G cell sleep and tunnel shutdown strategy. Follow up work is also pursuit in the Green 5G project that uses TM Forum Autonomous Networks technology to define a unified energy efficiency standard, and a methodology to deliver it across different business requirements, for instance in building energy saving capabilities for base stations.

Sustainable growth for enterprises with 5G and MEC operations Catalyst Project: Among various MEC and 5G connectivity use cases, the project addresses the challenge faced by electricity grids when operating legacy infrastructure by deploying the use case of smart energy management. The expected benefits are improved employee working experience, improved employee health & safety (EHS), improved overall operational efficiency, and a reduction in wasted energy and carbon emissions

3.4.5. Edge-cloud technologies in containerised data centres close to renewable energy sources

3.4.5.1. Introduction

Currently there are not widely known edge computing solutions based on modular/containerised datacentres connected directly to renewable energy sources.

In general, the idea of such edge computing comes from some limitations of standard data centres, which some of them are:

- High entry point from costs point of view
- Limited scalability
- Energy low efficiency and high CO2 emission [ERE-1], [ERE-2], [ERE-3]

From above limitations point of view, there are two main streams of possible solutions. One of them is investing in smaller, flexible solutions in the form of containerised datacentres.

Second is to optimize energy usage, paying attention to energy sources (ex. carbon power plant source vs. renewable energy source, like wind farm).

Edge computing seems to be the perfect approach to solve the problems mentioned above, as it helps to build competitive, smaller in size, and more cost-effective edge datacentres, and place them directly near energy sources, including renewable ones.

Chapters below describe the main components of the green edge processing use-cases separately because currently, as already mentioned, there is no one common, known solution combining all three main aspects of such approach, which are:

- Containerized data centres
- Edge cloud/computing management
- Direct connection to renewable energy sources

3.4.5.2. Containerised micro datacentres

Containerised datacentres are already known solutions produced and supported by a number of big vendors.

Currently they are used mainly in a few scenarios:

- When mobile datacentre needs to be moved from one place to another (ex. army)
- As an extension of standard data centre when building expansion is not possible
- As datacentre for small-medium companies (universities) with low-cost entry point

Shneider Electric is one of the players engaged in micro-datacentres delivery [ERE-4] from the smallest 6U solutions to bigger ones, called regional modular datacentres.

Other big player on the market is Dell [ERE-5], that delivers modular datacentres together with full set of equipment including computes.

Some other examples of such approach are Vertiv [ERE-6], Kstar [ERE-7] and Cisco [ERE-8] delivering professional and secure modular datacentres.

There is also concurrent approach to the modular and containerised datacentres building. There are small vendors, specialized in delivery of such solution on local markets. They compete with enterprise solutions by lower price, flexibility in constructing and local support availability on demand.

Such described containerised solutions can be used by means of edge cloud computing, building distributed network of small, mobile datacentres placed directly near green energy source.

3.4.5.3. Renewable energy sources in standard computing and cloud computing

Data centres are estimated to have been responsible for between 0.8% and 2% of the global consumption and 2.7% of EU energy use [ERE-3]. Growing energy consumptions force tech companies focus on energy usage efficiency and pay attention on energy source (green vs. carbon).

Green energy currently is mostly available to data centres from standard green and purchased as Green Energy Certificates. It is not a perfect solution as it just shows that such amount of energy was produced by renewable energy sources not consumed directly by data centre from green energy source [ERE-3].

Current trend is to go into PPA (Power Purchase Agreement) contracts, which helps to gain real access to energy produced locally in a given energy plant. This can help to realize real edge-cloud green computing scenario.

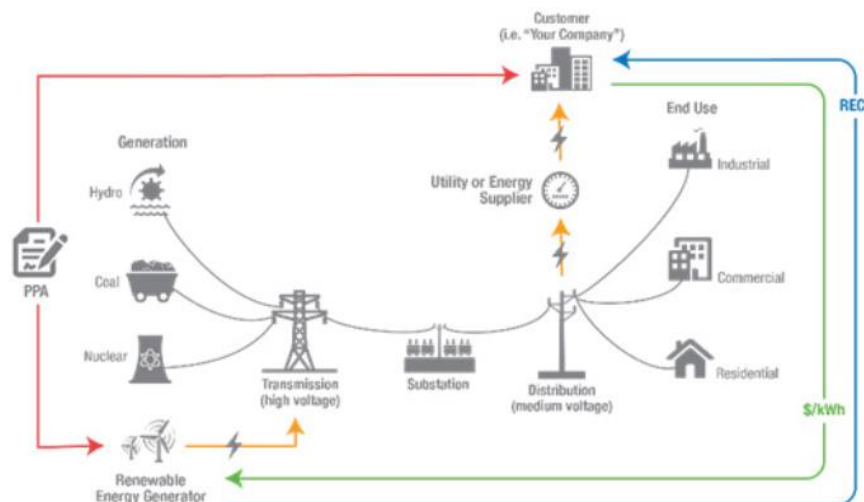


Figure 77: direct/physical power purchase agreement (PPA) [ERE-3]]

As some use cases show, it is possible to achieve high carbon-free energy consumption in computing. For example, Google Oklahoma scenario shows 96% of green energy in overall energy usage in data centre. From green edge computing scenario point of view, it is very important to have possibility to power edge datacentres directly from renewable energy sources by PPA contract scenarios, what let to achieve green, efficient processing solution.

3.4.5.4. Edge cloud management technologies for edge computing

Edge computing in edge datacentres requires some specific management solutions for edge cloud components provisioning, configuring, monitoring and processing distribution. Currently there are known a few initiatives which deliver some of the functionalities required by edge computing.

One of them is StarlingX [ERE-10], which helps to build multi-node edge cloud solutions, manage them, plan update/upgrade and installation cycles and finally monitor all infrastructure.

Other initiative from cloud computing is Airship, supported by OpenInfra foundation. The solution, aiming in telecommunication sector, allows operators to manage their infrastructure deployments and lifecycle.

Arkaino is one more edge cloud management solution [ERE-11] which helps realized distributed cloud computing scenarios.

Taking into account above existing examples, aerOS seems to be suitable solution which can help in building green edge processing use case, delivering one coherent environment from system, security, management and computing point of view.

3.4.5.5. Edge computing in combination with green energy sources

As mentioned in the introduction, currently there are no widely known, big solutions of a distributed edge computing in containerised datacentres connected directly to renewable energy sources. It is still an area of research to build such environment and check its usability, efficiency and level of being environmentally friendly.

aerOS itself with its planned functionalities can help in realizing the goal of real distributed edge computing.

4. Market analysis report

This section, differing radically from the previous in its scope, focuses on the potential market of aerOS instead of reviewing the scientific status of technologies. This market is understood as the niche of the global solution (meta operating system for orchestrating the IoT-edge-cloud computing continuum) and of the directly targeted sectors drawing from aerOS pilots (maritime ports, smart buildings, containerised data centres, manufacturing and construction, forestry and agricultural equipment)¹⁷.

4.1. aerOS market

4.1.1. Target Market

Digital transformation will represent an increasingly important aspect in the development of companies and there is an increasing awareness of its relevance, indeed 61% of CEOs declare that digital transformation is among their 3 top priorities¹⁸.

Europe needs products and solutions to accelerate digitization in areas with the most societal value. Today, 83% of EU SMEs do not use advanced cloud services. Europe needs to upskill and reskill the population and workforce of tomorrow by taking digital literacy to the next level. Today, over 42% Europeans do not have basic digital skills, while over 57% of companies are facing difficulties in finding ICT personnel. The time for doing more on upskilling, reskilling and inclusion is now.

Additionally, the increase in connected devices and in the volume of data coupled with evolving networks need such as lower latency and faster speed.

4.1.1.1. Cloud Computing market

Cloud has gained increasing importance in the development of new digital experiences, leveraging on drivers such as the pandemic and the increasing in digital services. Energy firms are leveraging cloud to better their customers' retail experiences, vehicle companies are providing new personalized services for customers' safety and entertainment, and the cloud has enabled new digital experiences like mobile payment systems for banks¹⁹.

This trend shows no hint of slowing down. Indeed, it is foreseen that by 2025, 85% of organizations will use cloud-native technologies that will become essential for their digital strategies and that over 95% of new digital workloads will be implemented on cloud-native platforms by 2025, up from 30% in 2021²⁰.

This growth will not be driven by a mere optimization of IT (e.g. IT cost optimization, risk reduction, core operations digitization), indeed, the majority of the growth will derive from innovation (e.g. faster production development, hyper-scalability)²¹. Indeed, \$770 billion of cloud's predicted value in run-rate EBITDA²² across Fortune 500 by 2030, equivalent to the 75% of the total predicted value of cloud will originate from innovation activities, while only \$430 billion will originate by rejuvenating activities²³.

Besides cloud will experience a strong growth in the next years, it will have to face some main concerns regarding security and access issues. Indeed, it represents the main concern about cloud for the 75% of

¹⁷ In this whole section, the references are inserted as footnotes, different than for Section 3. This is due to the fact that most of those references are websites that do not contain DOI and are not scientific. This has also been kept this way in order not to overburden the size of the reference section and to allow better readability of this (more operative) section.

¹⁸ [PwC, Time for trust - The trillion-dollar reasons to rethink blockchain, 2020](#)

¹⁹ [Gartner, Gartner Says Cloud Will Be the Centerpiece of New Digital Experiences, 2021](#)

²⁰ Ibidem

²¹ The value of the use cases in pioneer is not predictable yet so it has not been considered

²² Earnings before interest, taxes, depreciation, and amortization.

²³ [McKinsey & Company, Cloud's trillion-dollar prize is up for grabs, 2021](#)

enterprises. In this regard, the main challenges will concern infrastructure configuration, access, and insecure APIs.

In Europe Cloud computing is gaining increasing importance and, in 2021, 42% of EU enterprises used cloud computing, gaining 6 percentage points over the previous year (36%) and more than doubling compared to 2016 where the figure amounted to 19%. This value varies greatly in the different EU countries as, for instance, it amounts to 75% in Sweden and Finland while in Bulgaria and Romania less than one enterprise out of five uses cloud computing services, with a share of 13% and 14% respectively²⁴.

The fast-growing rates showed by the cloud computing market will require a large number of professionals trained in the field. Currently, skill shortage seems to be one of the biggest obstacles to the proper development of the market; in fact, skill mismatch is seen by 80% of cloud leaders as the biggest barrier to the cloud computing market²⁵.

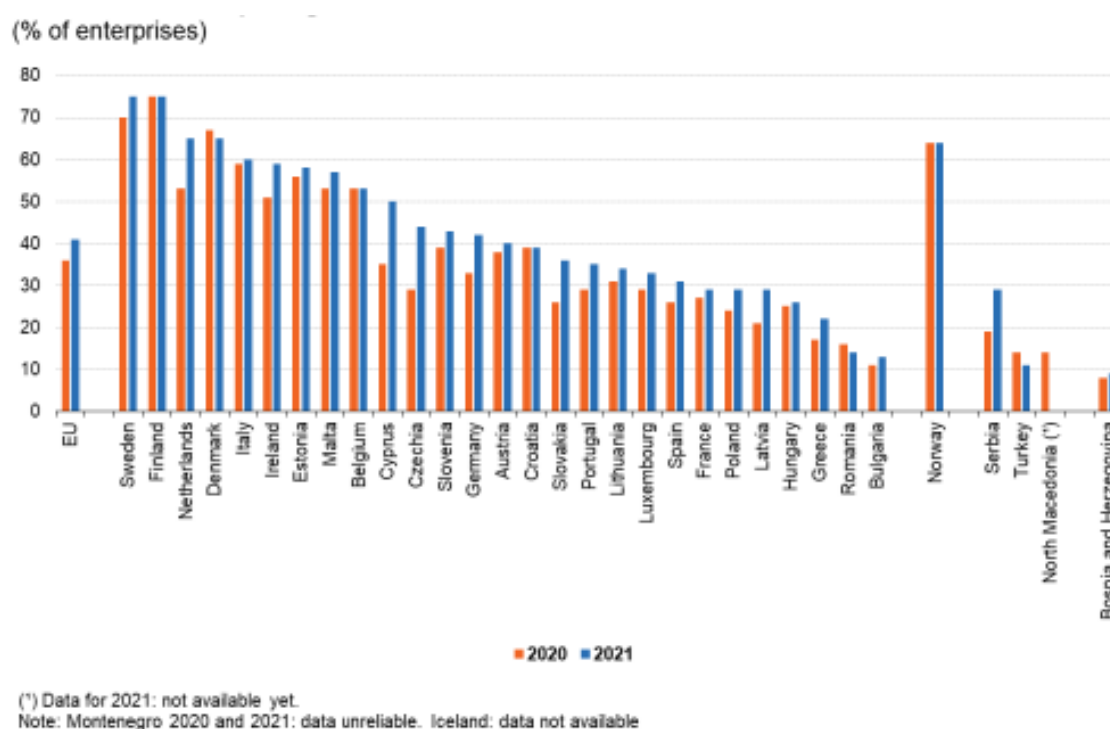


Figure 78: Use of cloud computing services, 2020 and 2021. Source: Eurostat, Cloud computing - statistics on the use by enterprises, 2021

There are several reasons why European enterprises use cloud computing services. Among them, the most relevant are e-mail (79%), storage (68%), office software (61%) and security software (59%) purposes. More interestingly, these businesses accessed more sophisticated end-user software programs via the cloud, including enterprise resource planning (24%), customer relationship management (27%) and financial/accounting (48%)²⁶.

4.1.1.2. Edge Computing market

According to Gartner, “the edge computing market provides the hardware, software and services to extend an agile digital enterprise to the edge, enabling lower latency, reduced data traffic, and semiautonomous computing”²⁷.

²⁴ Eurostat, Cloud computing - statistics on the use by enterprises, 2021

²⁵ Modis, Mission Possible: Tackling the cloud skills gap, heads on, 2021

²⁶ Ibidem

²⁷ Gartner, Market Guide for Edge Computing, 2022

Edge computing meets the increasing need to deal with the increasing data regulation, such as the General Data Protection Regulation (GDPR) in the European Union. Indeed, as of 2021, and with the increasing higher amount of data volume and velocity. Indeed, there are more than 60 countries around the world that have in place data protection localisation requirements. In this context, edge computing could be key to comply with these increasing requirements, locating computing infrastructure closer to the end user. Simultaneously, due to latency issues and the high costs related to move data, only less than 20% of data generated by enterprises are actually used²⁸.

These will lead to a marked increase of the edge computing market in the next few years, as it is expected the worldwide enterprise expenditure in edge computing will reach approximately \$ 250 billion in 2025, with an expected CAGR in the 2022-2025 period of around 10%. In addition, it is estimated that the percentage of servers shipped to companies that will be deployed at edge locations will grow from the 20% in 2019 to the 26% in 2024²⁹ and that the percentage of data generated by enterprises that will be created and processed at the edge will increase from 10% in 2018 to 75% in 2025³⁰.

The adoption of edge computing technologies is directly related to the current digital trends, among these, it is possible to appoint, for instance, the increasing adoption of IoT and digitalization and the need for a most efficient data management.

Currently, the edge computing market is rapidly expanding and evolving, the market features a high number of use cases, characterised by different vertical industries, requirements (e.g. low latency, high volume of data). This leads to the presence of many solutions that are first-of-a-kinds and highly customised and does not present a broader edge computing strategy. In this regard, Gartner has identified eight submarkets.

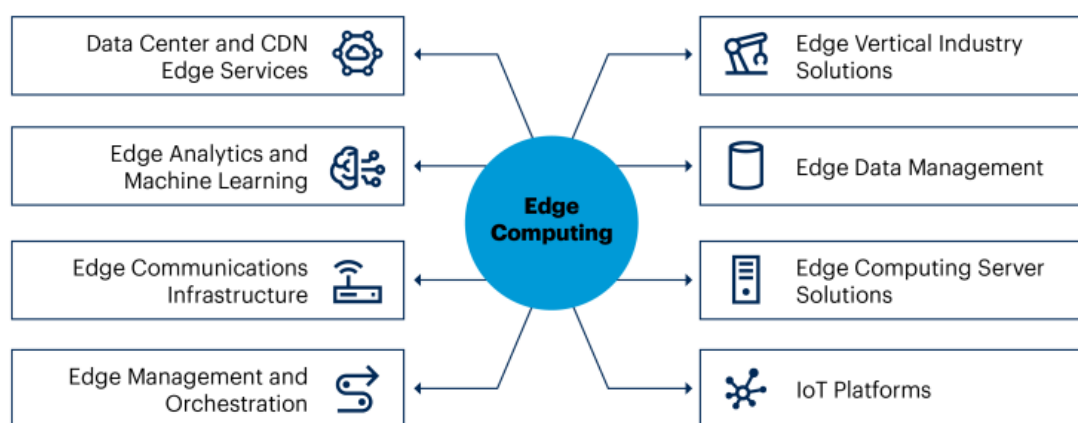


Figure 79. Edge Computing Submarkets. Source: Gartner, Market Guide for Edge Computing, 2022

4.1.2. Correlative Market

The edge and cloud computing market are closely related to other sectors. These are influenced by the edge and cloud computing markets and in turn influence these markets as drivers of development. These markets are, in particular, the IoT, AI, telecommunication and blockchain market.

These sectors show a strong inter-relation, and, in many cases, their combined use can enhance the opportunities offered by another technology. For instance, Artificial Intelligence technologies, in particular Machine Learning and Deep Learning, can be used in IoT applications to perform some edge computing tasks (e.g. distributed caching, quality of service optimization). At the same time, the combined use of multi-access edge computing (MEC) and AI can help in maximising the computing resources offered by edge computing³¹.

²⁸ [McKinsey & Company, McKinsey Technology Trends Outlook 2022, 2022](#)

²⁹ [Ibidem](#)

³⁰ [Gartner, What Edge Computing Means for Infrastructure and Operations Leaders, 2018](#)

³¹ [MDPI, Special Issue "Symmetry in Artificial Intelligence and Edge Computing", 2022](#)

4.1.2.1. IoT Market

IoT market is experiencing a strong growth, as in 2022 it has been valued at \$ 478.36 billion and is currently projected to reach \$ 2,465.26 billion by 2029, with a CAGR of 26.4% in the 2022-2029 timespan³².

Currently, in the world IoT market scenario, Europe is the third largest adopter of IoT after Asia-Pacific Region and North America: indeed, in 2019, Europe accounted for the 23% of global IoT spending, while Asia-Pacific for the 35.7% and North America for the 27.3%³³. The European spending experienced a steady increase, as in 2021 the European IoT spending amounted to \$ 202 billion³⁴.

European market has great room for growth as it has the characteristics to be applied to several verticals and it will translate in an increasing number of IoT devices. Indeed, while the number of active IoT devices currently on the market amounts to more than 10 billion, it is foreseen that by 2030 the devices will be 25.4 billion. This means that, by 2030, 23% of the devices will be located in Europe, while 26% in China and 24% in North America, making Europe an increasingly important market³⁵.

4.1.2.1.1. The European IoT Market

4.1.2.1.2. Current trends and drivers

The current IoT sector is influenced by three main topics and trends³⁶:

- the acceleration of innovation, driven by both the progresses made in the IoT specific sector (e.g., sensors, platforms and application technology) and the innovations in correlative markets, such as AI and edge computing, for technology and service providers (TSPs) that will supplement and replace operational technology (OT);
- the growth in number, variety and scope of the assets monitored through IoT sensors has led to a sharp increase in the volume of data that need to be processed. As a consequence, it is increasingly necessary to process these data closer to where they are originated;
- the need to provide an offer more aligned with the business instances will lead TSPs to put on the market more targeted products through a unique business which is focused on highly customised applications with the provision of embedded Artificial Intelligence and Machine Learning technologies.



IoT and cloud and edge computing³⁷

With the number of IoT devices doubling every five years³⁸, cloud and edge computing will have a fundamental role in the management of all the related data, both generated and to be processed. Indeed, increased use of the cloud has sped up the creation and implementation of scalable IoT applications and business models.

The increase in number of IoT devices influences the emergence of edge computing, as it requires to process data close to their generation sources, while edge computing is able to help in coping with their increasing volume and content that IoT devices generate. Indeed, thanks to what just mentioned, it would be possible to avoid the need to generate data in the cloud, thus allowing to make decisions right at the edge while at the same time keeping data storage and processing

³² [Fortune Business Insight, Internet of Things \(IoT\) Market to Witness 26.4% CAGR from 2022 to 2029; Oracle Corporation Launched Portable Server for Edge Computing to Expand Footfall, 2022](#)

³³ [CBI - Centre for the Promotion of Imports from developing countries, The European market potential for \(Industrial\) Internet of Things, 2022](#)

³⁴ [IDC, European IoT Spending to Exceed \\$200 Billion in 2021 as Companies Start Moving to the Next Stage of Recovery, According to IDC, 2021](#)

³⁵ [Centre for the Promotion of Imports from developing countries \(CBI\) - The Netherlands Ministry of Foreign Affairs, The European market potential for \(Industrial\) Internet of Things, 2022](#)

³⁶ [Gartner, Emerging Technologies and Trends Impact Radar: Internet of Things, 2021](#)

³⁷ This and the following box icons are made by [Freepik](#) from [www.flaticon.com](#)

³⁸ [Gartner, Gartner Predicts the Future of Cloud and Edge Infrastructure, 2021](#)

closer to the edge, maintaining only relevant and critical data to the cloud, avoiding round trip towards it. In this way, the management of future amount of data coming from IoT devices would become easier, also considering that it is estimated that more than 75.44 billion IoT devices will be in use by 2025, with a 500% increase compared to 2015³⁹.

The ability to process the collected data at the edge (on the device itself, before transmitting them over) and to enable significant bandwidth savings is therefore achievable thanks to both the decrease in cost and the increase in computing power of devices used in the IoT. In many circumstances, it also results in a better compliance with privacy requirements, since, as opposed to sending out raw data, data are gathered and encrypted on the device itself⁴⁰.

Furthermore, the edge IoT industry is predicted to experience considerable growth, as the total IoT market is expected to double between 2019 and 2024. Analysts anticipate the edge to expand by 35% yearly, as additional use cases and new technologies are developed, and the top edge use cases to represent up to 20% of the IoT market in 2024, compared to less than 10% of 2019⁴¹.

4.1.2.2. AI Market

Artificial Intelligence technology has strongly grown across the last years, making it easier and more affordable to implement, indeed, it has been estimated an improvement in training speed for AI models of 94.4% across the 2019-2021 timespan⁴². This goes hand in hand with a strong innovation effort reflected in a high number of patents registered in the period from 2015 to 2021, with a compound annual growth rate of 76.9%, as the number of patents filed in 2021, compared to 2015 is 30 times higher. This growth is led by the East Asia and Pacific region that accounts for the 62.14% of the total patent filings while the majority of the granted AI patents comes from the North America (56.96%) followed by East Asia and Pacific (31.09%) and Europe and Central Asia (11.27%) while all the other regions combined sum up to roughly 1% of total world granted patent⁴³.

This has been supported by strong private investments that, in 2021, touched \$ 93.5 billion.

Companies are increasingly adopting Artificial Intelligence (AI) technologies but with different rates among regions. The AI adoption rate by organisations in 2021 worldwide amounted to 56% with an increase of 6 percentage point compared to 2020. According to McKinsey survey, Europe AI adoption rate is slightly lower compared to the worldwide average, amounting to 51%. The areas with the highest adoption rates are the developed asia-pacific region (64%) and India (65%)⁴⁴.

Gartner⁴⁵ has identified the upcoming innovations in Artificial Intelligence. All these innovations fall into four main categories:

- **data-centric AI:** data-centric AI shifts the traditional focus of AI sector away from the improvement of AI models toward improving and enlarging the data used to train algorithms in order to get better outcomes from AI solutions. This will include synthetic data, knowledge graphs, data labeling and annotation;
- **model-centric AI:** despite the growing interest in improved data quality, future AI industry development trends have to include improved AI models. This will include focus on physics-informed AI, composite AI, causal AI, generative AI, foundation models and deep learning;
- **applications-centric AI:** edge AI and decision intelligence are at the core of this trend, the first aiming at embedding AI technologies at the IoT endpoints and the second enhancing decision making. It includes AI engineering, decision intelligence, operational AI systems, ModelOps, AI cloud services, smart robots, natural language processing (NLP), autonomous vehicles, intelligent applications and computer vision;

³⁹ [Atos, White paper Scientific Community, A 2021 perspective on edge computing, 2021](#)

⁴⁰ [CBI - Centre for the Promotion of Imports from developing countries, The European market potential for \(Industrial\) Internet of Things, 2022](#)

⁴¹ [Boston Consulting Group, The Battle at Computing's Edge, 2021](#)

⁴² [McKinsey & Company, McKinsey Technology Trends Outlook 2022, 2022](#)

⁴³ [Stanford University – Human-Centered Artificial Intelligence, Artificial Intelligence Index Report 2022, 2022](#)

⁴⁴ [McKinsey & Company, McKinsey Technology Trends Outlook 2022, 2022](#)

⁴⁵ [Gartner, What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle, 2022](#)

- **human-centric AI:** it involves technologies that can learn and collaborate with humans to enhance human capabilities. It includes AI trust, risk and security management (TRiSM), responsible AI, digital ethics, and AI maker and teaching kits.



AI and cloud and edge computing

Given the unlimited computing power of cloud computing, until now, it has been the natural fit for Artificial Intelligence services as it makes it possible to process massive amount of data. Nonetheless, in some cases, latency is fundamental (e.g. healthcare, transportation and robotics) or there are strict data protection requirements where personal data cannot be stored in a central repository. In addition, it may happen to have an area without network connection (e.g. underground, rural areas). In these cases, edge AI may be more adequate. Thanks to edge AI, Artificial Intelligence and Machine Learning are moved closer to the data generation and computation origin. In this way, edge computing will play a key role in the deployment of Artificial Intelligence technologies as it can help in ensuring the optimal conditions (e.g. low-latency and proper operation) needed, since some of these highly automated AI applications (e.g. manufacturing plant environment) rely on such network support. This will enhance the capabilities of AI-enabled application, lower operating costs and allow for a more efficient control over operations. It will, also, ensure data security and faster computing.

This technology may be beneficial for several applications, such as, for example, virtual assistants, that could train machine learning models through data stored on the device rather than in the cloud, autonomous vehicles, where edge AI could contribute to a better identification of road signs and to provide an higher level of safety and finally also automated optical inspection, making it easier and faster to detect manufacturing defects.

4.1.2.3. Telecommunication market

The telecommunication market is composed by telephone, telecommunication, and internet service providers. The market's constant considerable effort in innovation has been fostering the growth of the telecommunication market through the years. As of 2021, the telecommunication market was valued at \$ 2,642.14 billion and is expected to reach the value of \$ 2,866.61 billion in 2022 with a CAGR of 8.5%. The market is then expected to grow at a CAGR of 7.4% between 2021 and 2026, reaching by 2026 a value of \$ 3,818.36 billion⁴⁶.

In this context, 5G technology is one of the main priorities in the sector. The technology has the potential to reduce latency and offer an ultra-reliable coverage. In this sense, the investments in the 5G infrastructures will drive the market growth of communication service providers, with a predicted CAGR of 32.4% in the 2022-2030 period, reaching a value of \$ 95.88 billion by 2030. Applications such as ultra-high definition (UHD) videos, cloud-based AR/VR gaming, and HD video meetings will benefit from the high bandwidth connectivity offered by the 5G technology and it offers opportunities for a wide range of industries such as manufacturing, oil and gas, mining, and energy and utility⁴⁷.

Towards this innovation, Europe and North America show relevant differences. Indeed, while the European telecommunication market is highly fragmented showing a high number of telecom companies even in the smaller countries, the North America market is way more concentrated. This brings to a slower deployment of innovative technologies such as 5G indeed for example, while in the North America, the main operators, having a larger customer base have been able to offer 5G faster in the market. It is reflected in the number of 5G subscriptions in the two regions, indeed in Europe the number of subscriptions is much lower compared to the North America, with 5G accounting for the 6% of all subscriptions compared to the 20% of North America⁴⁸.

Despite the slow deployment of 5G subscriptions in Europe, it will be a key factor for the sector's growth in Europe, indeed, it is predicted that the 5G technology alone can have a strong impact on the European GDP, with a potential impact of € 113 billion annually in addition to a 2.4 million new jobs in 2025. The achievement of this result requires estimated € 150 billion investments in the sector. 5G in Europe will play an increasing relevant role and in 2025 it will account for two-thirds of total Telco revenues.

⁴⁶ [ReportLinker, Telecom Global Market Report 2022, 2022](#)

⁴⁷ [Bloomberg, 5G Infrastructure Market to be Worth \\$95.88 Billion by 2030: Grand View Research, Inc., 2022](#)

⁴⁸ [Reuters, Explainer: Why Europe's mobile telecom market is ripe for consolidation, 2022](#)

The table below shows the estimated revenues of the 5G market broken down for the different use case categories.

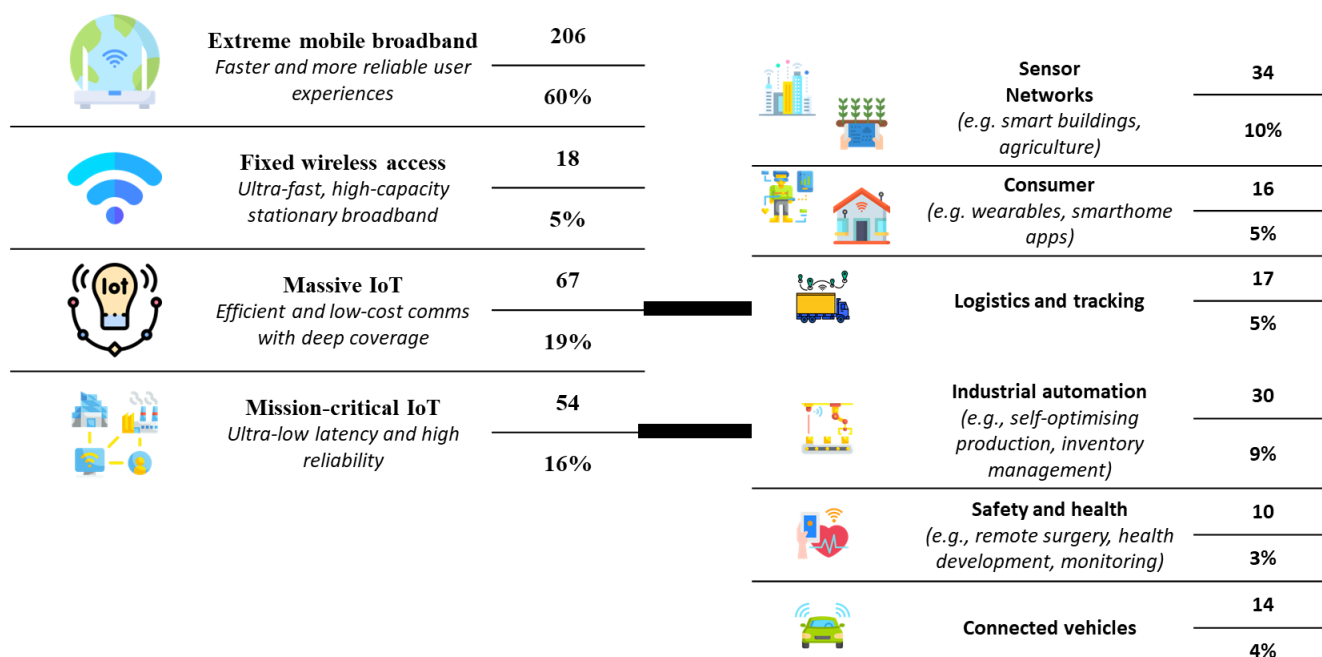


Figure 80: Estimated revenues in EU 2025 (€B). Source 1: European Telecommunications Network Operators' Association (Etno), Connectivity & Beyond: How Telcos Can Accelerate a Digital Future for All, 2021

Icons by Flaticon.com

(% share of use case category of total revenues)



Telecommunication and cloud and edge computing

The telecommunications market will be significantly impacted by cloud and edge computing. The key impact is an expansion of revenue sources from technologies like multi-access edge computing (MEC), given the telecom industry's position as the primary owner of the networking infrastructure needed for distributed computing⁴⁹.

In addition, edge computing gives telecommunication companies the possibility to speeding up mobile applications improving their performance, as well as increasing their efficiency reducing, for example, network congestions⁵⁰ [...]

Furthermore, telecommunication innovation such as 5G will be one of the main drivers of edge computing and IoT market growth. Indeed, these technologies have a higher spectrum efficiency and provide high bandwidth and low-latency, that represent key aspects for edge computing and represent an enabler for IoT technology.

4.1.2.4. Blockchain market

The global blockchain market is experiencing a strong growth. Indeed, with a predicted CAGR of 56.3% over the projection period of 2022-2029, the worldwide blockchain market is expected to increase from \$7.18 billion in 2022 to \$163.83 billion by 2029. The market has been heavily impacted by Covid-19 pandemic experiencing a strong decrease in demand compared to pre-pandemic levels, with a decline of 52.4% in 2020 compared to 2019⁵¹.

⁴⁹ [McKinsey & Company, McKinsey Technology Trends Outlook 2022 - Cloud and edge computing, 2022](#)

⁵⁰ [Bloomberg, Burgeoning Adoption of Internet of Things \(IoT\) to Steer Global Edge Computing Market Past US\\$ 69 Bn through 2032, 2022](#)

⁵¹ [Fortune business insights, Blockchain market size, share & Covid-19 impact analysis, 2022](#)

It is expected that the blockchain-derived business value will increase quickly, with a forecasted value of \$ 176 billion by 2025 and \$ 3.1 trillion by 2030⁵².

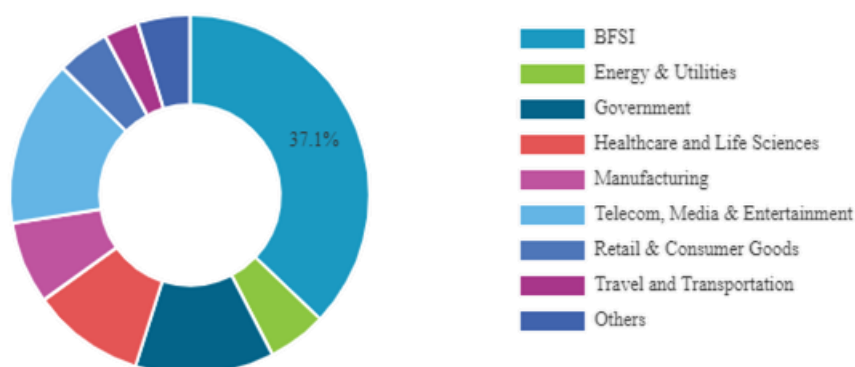


Figure 81 - Global Blockchain Market Share, by industry, 2021. Source: Fortune business insights, Blockchain market size, share & Covid-19 impact analysis, 2022

As displayed in the figure above, the Banking, Financial Services and Insurance (BFSI) market holds the largest market share as blockchain allows to optimize company processes and decrease operational costs. Besides the importance of the BFSI market, the sector with the highest anticipated CAGR is retail and consumer goods.

The blockchain market is expected to have a strong impact also on the global GDP⁵³, increasing from an impact of \$ 66 billion in 2021 to \$ 422 billion in 2025, reaching a value of \$ 1.76 trillion in 2030 accounting for the 1.4% of global GDP⁵⁴.

At the European level the ecosystem and the regulatory maturity of the blockchain sector is heterogeneous among the European countries. Indeed, while Cyprus, Estonia and Malta have a strong presence of a local business/startup ecosystem, a valuable blockchain-related formal education and academic research initiatives and a developed user-driven communities around blockchain or virtual assets and have a defined regulatory system in the blockchain market. On the opposite side of the spectrum Belgium, Bulgaria, Croatia, the Czech Republic, Greece, Hungary, Romania, and Slovakia show a low level of development of the market and have not developed a regulatory framework⁵⁵.



Blockchain and edge computing

With an infrastructure to store and validate transactions, edge computing could help blockchain overcome its difficulties.

Moreover, the manner that network designs are constructed now is a component that causes delays in blockchain networks. Similar to how traffic flows in cloud computing, data must travel through the entire network and back in order for blockchain nodes to connect with one another. A server-to-server data flow would be enabled via an edge computing network, eliminating the requirement for data to pass through the core network.

4.1.3. Market Size and growth

Indeed, it is deemed that edge computing will increasingly become an operational necessity for businesses. In this perspective, it is estimated that in 2025 the projected worldwide spending in edge computing will reach around \$ 250 billion⁵⁶.

⁵² [Gartner, Digital Disruption Profile: Blockchain's Radical Promise Spans Business and Society, 2022](#)

⁵³ It refers to GDP (in US\$, 2019 prices) which is the net additional value created by blockchain.

⁵⁴ [PwC, Time for trust - The trillion-dollar reasons to rethink blockchain, 2020](#)

⁵⁵ [EU Blockchain Observatory & Forum, EU Blockchain ecosystem developments, 2020](#)

⁵⁶ [McKinsey & Company, McKinsey Technology Trends Outlook 2022 - Cloud and edge computing, 2022](#)

Enterprises will increasingly adopt edge computing solution, indeed, it is estimated that by 2023 over 50% of new enterprise IT infrastructure deployed will be at the edge rather than corporate datacentre. It will be a huge increase from the 10% of today⁵⁷.

Edge cloud computing represents a fundamental market for the “Fourth Industrial Revolution” that will play an increasingly important role to support the deployment of the Internet of Things (IoT) ecosystem, the global sharing economy and the increase of zero marginal cost manufacturing.

To meet the rising demand for devices and edge infrastructure, enormous infrastructure expenditures are required. It is predicted that between 2019 and 2028, edge computing infrastructure and new and replacement IT server equipment would require cumulative capital expenditures of up to \$800 billion USD⁵⁸.

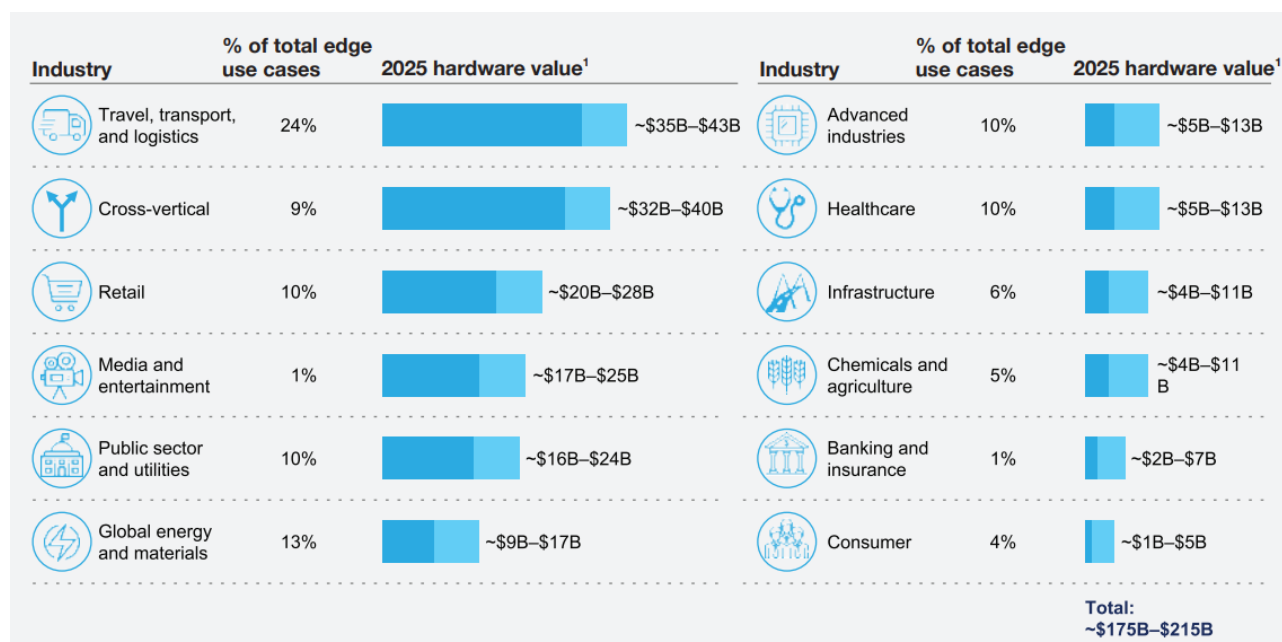


Figure 82. Edge computing potential value by 2025. Source: McKinsey & Company (JM Chabas, C. Gnanasambandam, S. Gupte, and M. Mahdavian), *New demand, new markets: what edge computing means for hardware companies*, 2018

Note: Hardware value includes opportunity across the tech stack (ie, the sensor, on-device firmware, storage, and processor) and for a use case across the value chain (ie, including edge computers at different points of architecture).

Processing data locally besides representing a technical advantage, avoiding problems such as latency, represents an economical convenience for companies⁵⁹.

Currently, edge computing represents only a marginal solution, indeed, 10% of enterprise-generated data is created and processed outside a traditional centralized data center or cloud. The

4.1.4. Market Trends

4.1.4.1. Cloud computing market trends

Cloud computing will evolve considerably and will play a fundamental role in supporting the delivering of emerging technological innovations, e.g., in the field of Artificial Intelligence, Internet of Things and edge computing. Furthermore, cloud computing will become an essential asset for companies for almost any digital business initiative, with them using industry cloud platforms to accelerate digital innovations that are expected to pass from the 5% in 2022 to the 50% in 2027. The increasing adoption of cloud will be more and more

⁵⁷ [Atos, White paper Scientific Community, A 2021 perspective on edge computing, 2021](#)

⁵⁸

⁵⁹ [Equinix and Azure, an examination of the global impact and future of edge computing, 2020](#)

pervasive, until representing the main style of computing by 2027: it is estimated that, by then, cloud computing will be considered indispensable, or at least heavily impactful, on company organisation.⁶⁰

This trajectory of growth will bring cloud computing technology to move from simple technology enabler to a real business disruptor. The image below presents the predicted cloud computing projection from the current situation (phase 1 and 2) to 2027 (phase 3 and 4).

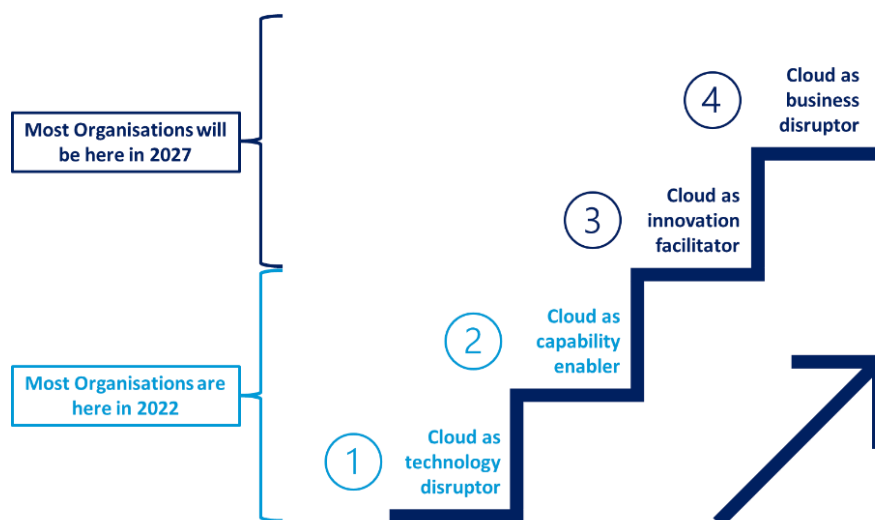


Figure 83: Cloud computing in 2027. Source: Own elaboration based on Gartner, *The Future of Cloud Computing in 2027: From Technology to Business Innovation*, 2022

The transition to cloud as a business disruptor will be rather rapid. While most businesses are currently in phases 1 or 2, both of which represent a still-relatively underdeveloped stage of the market (where cloud computing cannot yet provide all its potential opportunities), in just 5 years the majority will have moved into phases 3 or 4, which represent the most mature stages of the market (where cloud computing can manifest its full potential).

Currently, in phase 1, cloud computing is acquiring a more and more central role in enterprise investments decisions: with the advent of cloud-based technology, businesses may now see IT spending as an ongoing operational expense (OpEx), instead of an upfront capital expense (CapEx). Several companies are in a more advanced stage (phase 2), where new capabilities are available thanks to cloud, allowing the deployment of some AI-infused processes. Yet, this phase still suffers from economical burdens. In the future, most of the companies will shift to phase 3, where cloud starts enabling the widespread of platform business model: this phase represents the stage where technology starts becoming a fundamental component of enterprises business model. Finally, when the market will be mature, cloud computing will reach phase 4, where cloud will propose itself as the disrupting leader in the market. In addition to this, in the next few years, environmental aspects will be one of the most important factors taken into account in purchasing choices for hyperscale cloud services, ranking among the three most important by 2025.⁶¹

4.1.4.2. Edge computing market trends

In a similar way of what happens to cloud computing, the edge computing market is also set to experience major changes in the applications of the technology as it matures. To date, it is still in its early stages, being more concerned with the business results than with the IT architecture: this is mostly due to the fact that edge computing solutions are part of business or operational enhancements (such as enhancing the customer experience or automating factories) that fall within the purview of business people rather than IT specialists. Nowadays, the growth of the edge computing industry is influenced by applications in specific industries or by the emergence of particular needs, such as low latency or data processing of a big volume.

The development of the edge computing market will be divided into three main phases: the first phase of edge computing will lead to the creation of similar solutions for businesses operating in the targeted verticals. Thus,

⁶⁰ [Gartner, The Future of Cloud Computing in 2027: From Technology to Business Innovation, 2022](#)

⁶¹ Ibidem

in the second stage of the edge computing business, specialised solutions will be developed for particular sectors.

While the market is currently intensely focused on the creation of particular use cases, it is anticipated that within the next five to ten years it will gradually shift away from this approach and become more cantered on edge computing strategies and architectures.

Finally, as the market reaches its full maturity, horizontal solutions and the submarkets mentioned in the previous paragraph *Edge Computing market* will consolidate: this will be the third and final stage and it will mark the full development of the edge computing market, when its technological entire potential and benefits will be made available to the economy⁶².

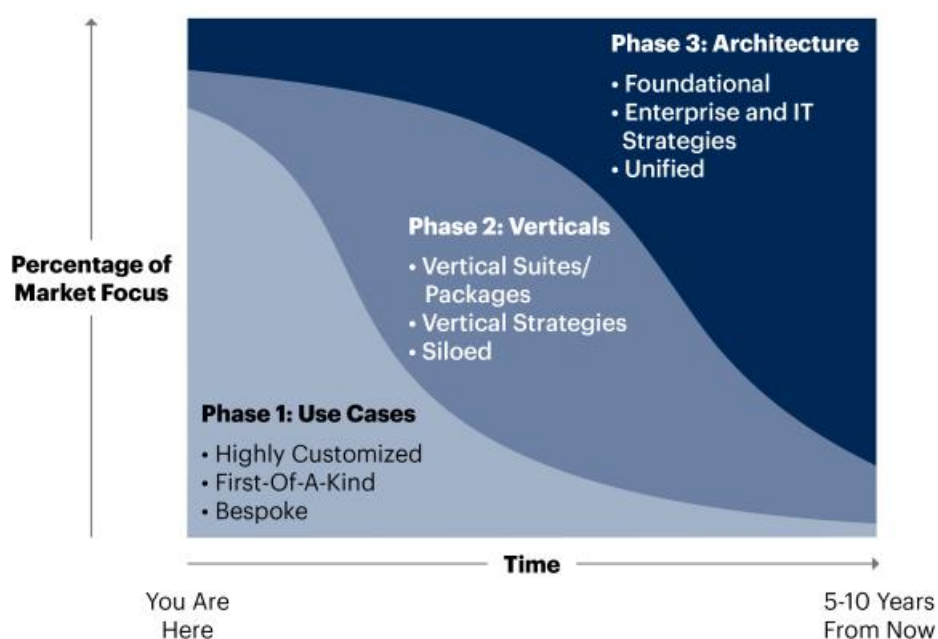


Figure 84. Phases of Edge Computing Market Focus. Source: Gartner, Market Guide for Edge Computing, 2022

⁶² [Gartner, Market Guide for Edge Computing, 2022](#)

4.2. Influencers market factors



Political

- Lack of a common political agenda for the adoption of edge to cloud solutions at EU level.
- Presence of tax incentives for the adoption of edge to cloud systems in EU Countries.
- Political choices regarding regulation have posed legal restrictions in several EU countries against the connection to renewable energy sources.



Economic

- High investment costs associated with the deployment of aerOS solution.
- Positive impact on correlated market and on the supply chain.
- Potential entry barriers to the market.
- Potential reluctances to the adoption of a solution derived from a EU funded project.
- Increase in European competitiveness.



Social

- Availability of skilled professionals.
- Social request for digital transformation.
- Improvement of working conditions.
- Privacy and security concerns.



Technological

- Strong interrelation with other technological solution and mutual enhancement.
- Difficulties to overcome reluctances and make disruptive technologies accepted.
- Marked improvement of current technologies.



Legal

- Privacy and security as increasing concerns for the edge to cloud computing.
- Lack of legal regulations related to cloud and edge computing.
- Uncertainty regarding future ineluctable regulatory framework.
- Lack of homogeneity among EU countries regarding regulation.



Environmental

- Data centres are extremely energy-intensive.
- Environmental factors are increasingly relevant in executives and customers decisions.
- The growth of the edge and cloud computing market, coupled with the growth of the correlative markets (e.g. IoT) will require an higher number of data centres and infrastructure.

These icons are made by [Freepik](https://www.flaticon.com) from www.flaticon.com

4.2.1. Political

It is generally acknowledged that the presence in the EU countries of economic regulations and tax incentives for the adoption of Edge to Cloud systems (or, in reverse, the lack of it) represents a significant discriminating factor from the perspective of transversal digital transformation across the EU territory. The decisions regarding tax incentives are currently left to the discretion of the governments of individual Member States, without a common general vision from the point of view of the European Institutions, thus creating a fragmented situation.

A general uniformity of European political agendas and horizons (both on a national Member State level and on a European Union one) with regards to the drive for adoption of Edge to Cloud solutions would represent a significant change in terms of political contribution to modernisation of the European society. An effective driving factor could come from the incentives provided by the national governments, and occasionally their direct commissioning of initiatives to generate novel and inventive technological products.

In fact, significant hurdle for several less digitalised countries is represented by an insufficient national internet infrastructure whose improvement requires intervention from the government.

Some players are facing some major obstacles in implementing some industrial application of edge to cloud technologies. For instance, in terms of Green Edge Processing, that is to say Edge-Cloud processing related to renewable energy sources focusing on sustainability and cost reduction, the political choices regarding regulation represents the main barrier from the point of view of the market. In fact, technologies like aerOS solution are usually developed not only for the internal use of individual Member States, but to be spread across the whole EU territory creating a European technological standardisation: nonetheless, a crucial issue is constituted by the fact that in several countries (e.g., Poland) the presence of legal restrictions against the connection to renewable energy sources persists despite the obvious environmental instances coming from all corners of the continent.

4.2.2. Economical

The deployment of an innovative solution such as a platform-agnostic meta operating system for the IoT edge-cloud continuum will require high costs related to the substantial and time-consuming study and implementation effort in order to identify, for example, functional or non-functional, technical and business requirements and to deliver the best architectural design and functional and technical specification.

A platform-agnostic meta operating system for the IoT edge-cloud continuum can provide positive effects to the Economy, both supporting the deployment of other technological solutions working as a catalyser and providing secondary industry impacts, positively influencing the whole supply chain.

Since there are no other competitor solution in the market, aerOS solution would benefit from the first-mover advantage. At the same time, it risks being perceived by the industry as a disruptive solution that breaks with current equipment, techniques and processes, requiring gigantic adoption efforts even though it is in the essence of aerOS to be built on top of current infrastructure to bring added value. Furthermore, the innovativeness could make it encounter a number of entry barriers, such as:

- the technological lock-in of digital solutions, where processes in an enterprise are locked to a specific solution, making the transition to a new one more difficult, even if it were more effective. It is combined with technological inertia and reluctance to take risks;
- concerns about user understanding and acceptance and the learning curve of a new and disruptive solution like aerOS. In particular, for smaller economic realities, a barrier is represented by the lack of technical abilities of the team in dealing with new disruptive technologies. This kind of technical barrier is less perceived as present in bigger terminals;
- potential lack of interoperability and user friendliness as significant barriers. Therefore, the reluctance against digital solutions and technological inertia, together with lack of heterogeneity of data, that makes them difficult to manage. In fact, data management algorithm should be easy to be managed by the end user and be customised on the basis of the situation in which it has to be used;
- slowdowns and obstacles in communication with different departments of large entities, with too long procedures that discourage investors;

- concerns related to the open-access nature of the solution leading to privacy and security concerns regarding open platforms and the presence of numerous business strategies in many sectors that do not include open solutions;
- the IoT edge-cloud computing sector is highly driven by de-facto standards (e.g., microservices, cloud-native technologies) and the uptake of solutions is influenced by vendor reference. AerOS will thus liaise with standardisation entities (membership by partners) and will rely on the adoption of aerOS by relevant project partners and stakeholders (JD, SIEMENS, TID, TTC – and its third linked party TCAG, ERICSSON, CF, ELECT etc.)

AerOs market is also influenced by potential barriers to the adoption of an EU funded project solution:

- market reluctance against products or services delivered from European projects, based on concerns about the feasibility, functionality and applicability of the solutions beyond the limited and controlled pilots where they were tested and developed. The use case scenarios are perceived as non-completely realistic because of their constraint and limited environments: for example, the security issue, which is pivotal in real life, is perceived by potential customers as too simplified in the laboratory and therefore not properly dealt with;
- since the prototypes from funded R&D projects are usually matured for sale after the project within 3-7 years depending on the application domain, the adoption issue does not concern encountering internal company reluctance, but further investments and time needed to advance the prototypes to a product level, and rapidly changing customer requirements;
- in certain fields such as ports, the market reluctance has not been clearly associated either with the “limited” European project development of the solution or with the general technological inertia in front of new and different digital solutions. These new solutions are often not acknowledged in their disruptive significance, and therefore not considered necessary and worth of such a big investment as the one they require to be integrated in the customer own system.

AerOs provides a technological solution that, offering resilience and flexibility in implementing faster responses to industrial requirements and unplanned events, can serve several application such as for the devices for the smart cities, including cameras analysing traffic and controlling traffic lights (need for them to act in a coordinated matter), smart monitoring and remote control of public infrastructures, support to AI technologies in all the phases of the manufacturing process, also considering the supply chain management, applications for the digitalisation and automation in ports.

It must be pointed out that the development of a platform-agnostic meta operating system for the IoT edge-cloud continuum could also have a significant meaning for the increase of European global competitiveness. Indeed, since Europe has already lost race for cloud technologies against America or Asia, the race for creating open tools and standardising IoT ads and related deployments represents a fundamental opportunity for European Countries to stay competitive if they reach the goal on time, also with ambitious solutions such as aerOS. The Edge to Cloud continuum system is pivotal for European non-dependence, sovereignty, and for a stronger position of European industry in the global market (including the whole value chain, e.g., technological components, systems, and so on). Furthermore, it is extremely important for Europe to invest on the Cloud continuum in order to optimise either the performance of existing services and products or to be the triggering point that will help the developers to introduce into the market innovative solutions, new services and new products. The Cloud Continuum will be even more revolutionary than the Cloud Computing was when introduced.

4.2.3. Social

The deployment of an innovative solution such as a platform-agnostic meta operating system for the IoT edge-cloud continuum will require high costs related to the substantial and time-consuming study and implementation effort in order to identify, for example, functional or non-functional, technical and business requirements and to deliver the best architectural design and functional and technical specification.

A platform-agnostic meta operating system for the IoT edge-cloud continuum can provide positive effects to the Economy, both supporting the deployment of other technological solutions working as a catalyser and providing secondary industry impacts, positively influencing the whole supply chain.

Since there are no other competitor solution in the market, aerOS solution would benefit from the first-mover advantage. At the same time, it risks being perceived by the industry as a disruptive solution that breaks with current equipment, techniques and processes, requiring gigantic adoption efforts even though it is in the essence of aerOS to be built on top of current infrastructure to bring added value. Furthermore, the innovativeness could make it encounter a number of entry barriers, such as:

- the technological lock-in of digital solutions, where processes in an enterprise are locked to a specific solution, making the transition to a new one more difficult, even if it were more effective. It is combined with technological inertia and reluctance to take risks;
- concerns about user understanding and acceptance and the learning curve of a new and disruptive solution like aerOS. In particular, for smaller economic realities, a barrier is represented by the lack of technical abilities of the team in dealing with new disruptive technologies. This kind of technical barrier is less perceived as present in bigger terminals;
- potential lack of interoperability and user friendliness as significant barriers. Therefore, the reluctance against digital solutions and technological inertia, together with lack of heterogeneity of data, that makes them difficult to manage. In fact, data management algorithm should be easy to be managed by the end user and be customised on the basis of the situation in which it has to be used;
- slowdowns and obstacles in communication with different departments of large entities, with too long procedures that discourage investors;
- concerns related to the open-access nature of the solution leading to privacy and security concerns regarding open platforms and the presence of numerous business strategies in many sectors that do not include open solutions;
- the IoT edge-cloud computing sector is highly driven by de-facto standards (e.g., microservices, cloud-native technologies) and the uptake of solutions is influenced by vendor reference. AerOS will thus liaise with standardisation entities (membership by partners) and will rely on the adoption of aerOS by relevant project partners and stakeholders (JD, SIEMENS, TID, TTC – and its third linked party TCAG, ERICSSON, CF, ELECT etc.)

AerOs market is also influenced by potential barriers to the adoption of an EU funded project solution:

- market reluctance against products or services delivered from European projects, based on concerns about the feasibility, functionality and applicability of the solutions beyond the limited and controlled pilots where they were tested and developed. The use case scenarios are perceived as non-completely realistic because of their constraint and limited environments: for example, the security issue, which is pivotal in real life, is perceived by potential customers as too simplified in the laboratory and therefore not properly dealt with;
- since the prototypes from funded R&D projects are usually matured for sale after the project within 3-7 years depending on the application domain, the adoption issue does not concern encountering internal company reluctance, but further investments and time needed to advance the prototypes to a product level, and rapidly changing customer requirements;
- in certain fields such as ports, the market reluctance has not been clearly associated either with the “limited” European project development of the solution or with the general technological inertia in front of new and different digital solutions. These new solutions are often not acknowledged in their disruptive significance, and therefore not considered necessary and worth of such a big investment as the one they require to be integrated in the customer own system.

AerOs provides a technological solution that, offering resilience and flexibility in implementing faster responses to industrial requirements and unplanned events, can serve several application such as for the devices for the smart cities, including cameras analysing traffic and controlling traffic lights (need for them to act in a coordinated matter), smart monitoring and remote control of public infrastructures, support to AI technologies in all the phases of the manufacturing process, also considering the supply chain management, applications for the digitalisation and automation in ports.

It must be pointed out that the development of a platform-agnostic meta operating system for the IoT edge-cloud continuum could also have a significant meaning for the increase of European global competitiveness. Indeed, since Europe has already lost race for cloud technologies against America or Asia, the race for creating

open tools and standardising IoT ads and related deployments represents a fundamental opportunity for European Countries to stay competitive if they reach the goal on time, also with ambitious solutions such as aerOS. The Edge to Cloud continuum system is pivotal for European non-dependence, sovereignty, and for a stronger position of European industry in the global market (including the whole value chain, e.g., technological components, systems, and so on). Furthermore, it is extremely important for Europe to invest on the Cloud continuum in order to optimise either the performance of existing services and products or to be the triggering point that will help the developers to introduce into the market innovative solutions, new services and new products. The Cloud Continuum will be even more revolutionary than the Cloud Computing was when introduced.

4.2.4. Technological

As discussed in more detail in the previous sections (cfr. [Correlative Markets](#)), cloud and edge computing continuum solutions are influenced by, and influence in return, the growth of other technology solutions (e.g., Artificial Intelligence, Internet of Things). So, in the next coming years the actual deployment of Edge-To-Cloud solutions will produce impacts on the development of other emerging technologies. More specifically:

- the number of IoT devices will experience a marked increase in the next few years and the Edge-To-Cloud framework will be essential to ensure an effective data storage and processing;
- artificial Intelligence requires to process a huge amount of data, to have a low latency and to accomplish high data security standards. In this regard the Edge-To-Cloud technological solution will enhance Artificial Intelligence capabilities;
- the Edge-to-Cloud framework represent a key aspect also for the growth of the telecommunication market as it allows telcos to quickly address changing needs from customers as well as to collaborate with Edge-to-Cloud providers as to provide innovative solutions to the market. Edge computing will also be key in the delivery of 5G technologies, in the data transfer and digital transformation of companies;
- the blockchain market is strictly related to the Edge-to-Cloud computing as the former offers a solid base for the deployment of some solutions related to the Edge-to-Cloud continuum, such as the Cloud of Things (CoT) and the latter can successfully support the development of blockchain, for example, storing transactions.

AerOS will include ambitious cybersecurity-related features that usually take time to be accepted and adopted and will rely on open-source technologies that might imply un-tested functionalities. To cope with this, aerOS will perform serious State of the Art and market analysis (T2.1) that will be sustained throughout the project (T6.4) to ensure that technological evolution speed of related areas (that is another barrier itself due to the high specialisation rate) is under continuous observance. AerOS will develop novel smart services and orchestration prepared for potential (at large scale uptake) encounter of connectivity, topology and configuration scenarios not initially considered.

AerOS, as a platform-agnostic meta operating system for the IoT edge-cloud continuum, will allow to overcome some issues related to the current cloud software framework, such as reducing latency or delay that may slow data processing and result in a reduced customer experience. It will also improve data compatibility, portability, automation, and interoperability across any component federation and the IoT edge-cloud continuum.

4.2.5. Legal

aerOS will make it possible to address legal requirements regarding data security. In fact, storing critical data away from centralised servers at edge sites helps limiting access and reducing risks in case of a significant attack. Additionally, as there are more edge locations, there are more places where malicious actors may attack: aerOS will take the necessary precautions in this regard to guarantee the highest level of protection against vulnerabilities.

Currently, there are several ethical and legal barriers for the adoption of a solution concerning the IoT edge-cloud continuum:

- ethical and legal issue are commonly due to the uncertainty of future regulations and future mandates (especially concerning IPR, confidentiality and privacy issue related to data management, need for anonymisation of data, personal data protection and commercial confidence) makes stakeholders and entities reluctant to invest. This prevents technologies from advancing at the rhythm they should. Furthermore, there is the need to add some rules inside the firewall for the data coming from external sources, with all the relative costs;
- relinquishing the control to third party services and storing data and other applications in the cloud could provoke legal and regulatory concerns. Efficient ways would have to be determined in order to keep the organisation in step with compliance requirements, such as anonymisation tools and GDPR mechanics. Indeed, GDPR compliance and compliance with National legal legislation might be different with the cloud provider. Operational applications and services will include personal information that needs to be anonymised;
- adoption hurdles due to the compliance with internal corporate IT policies for any Cloud solution;
- for Universities and Research Entities, the main potential hurdle could be represented by internal policies about opening networks to the WAN;
- the aspects of integrity and Company Use security are technologically challenging and might hurt significantly the interests of individuals and of larger parts of society if no appropriate technology is implemented in new production systems;
- although some extra EU countries are GDPR compliant (e.g., Norway, U.K., Iceland, Lichtenstein, Switzerland), others are not, and legal issues could arise if the cloud servers are located there. Furthermore, for companies that evaluate the “ethical score” of their subcontractors in terms of their ESG policy, this evaluation would be more difficult for cloud services, as these may be anywhere in the world;
- in several countries (e.g., Poland), legal restrictions against the connection to renewable energy sources are still present.

In addition to this, the players in the sector have to consider the uncertainty regarding future regulatory framework. Indeed, edge computing sector is a fast-changing sector and laws may change in the timespan of the project.

Even though the aforementioned ethical and legal barriers may represent a concern for the development of a technology regarding the IoT edge-cloud continuum, there are several solutions that could be applied:

- anonymisation and pseudonymisation policies;
- GDPR compliance: definition of data governance structures, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights concerning access to and processing of the data as indicated in EU Data Act⁶³;
- the Privacy Impact Assessment (PIA) improvement for organisations and the development of a single data EU market where privacy and data protection, as well as competition law, are respected and the rules for access and use of data are fair, practical and clear;
- pooling European data in key sectors, with common and interoperable data spaces;
- It is generally regarded as easier to achieve systems security when situated on the Cloud, and also to implement Disaster Recovery Systems and Redundancy Systems, as well as scalable and virtually unlimited retention period of records.

Currently, there are legal differences in the adoption of cloud or on premises systems:

- the current legal scenario does not regulate the data sovereignty issues that act as barriers to adopt external cloud-based solutions. There are no legal proofs that the data are not going to be transferred or re-used outside the country of origin, with or without the consent of the data owner.

⁶³ European Commission, proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (COM(2022) 68 final), 2022 - [1_EN_ACT_part1_v8.docx \(europa.eu\)](#)

- it could be considered legally easier to adopt Cloud-based systems, as some of the security obligations for some specific sectors (e.g., ports) originate from the National Computer Security Incident Response Team regulations.

Other factors influencing the legal context derives from the fact that different geographies have different regulations and different approaches to the topic “Privacy vs. Innovation”, that is to say the precarious balance between the need for innovation and privacy concerns when dealing with previously unexplored technological territories: this creates a serious reluctance to invest.

In the IoT edge-cloud sector, solutions are often rushed to market without proper security, transparency and privacy testing with ill-defined data policies. AerOS will embed data governance procedures (as part of the architecture) and will make frugal explainable AI one of its flagships to improve adoption in the mid- and long-term. Ethical and legal frameworks will be continuously observed through T1.3. The EU consists of multiple countries with different cultural backgrounds, which also apply in terms of digital technologies adoption: the fact of including partners from 11 countries in the Consortium will offset any potential bias. Finally, while the Consortium has all expertise and skills to set up aerOS towards deployment in real cases (IoT, Cloud, Edge, Security, Trust, Data technologies), that might not be the case outside of it. AerOS will deliver an easy-to-use solution allowing further installation, fostering a wider adoption of the project solution at a pan-European and international level.

4.2.6. Environmental

Data centres are particularly energy-intensive, as they are responsible for 4% of the world power consumption and 1% of its greenhouse gas emissions. Operators of IT infrastructure must continually adapt to the difficult demands on energy supply brought on by the gradually growing volume of digital transactions. Energy consumption will increase as a result of the growing need for edge/cloud computing, meaning a higher quantity of crucial infrastructure, number of devices, data centres, and associated energy needs, as well as the impact on correlative markets (e.g., IoT). This could embed high energy and CO₂ footprint. Indeed, more data storage and management capabilities are always needed, since the amount of stored data is predicted to increase by 5.3 times by 2025 compared to 2018⁶⁴. For this reason, data centres will increasingly rely on green IT strategies (e.g., electricity from renewable sources and energy-efficient cooling techniques). In this context, since more data is processed locally rather than being transferred over the network, edge computing in this scenario will substantially decrease total energy consumption⁶⁵.

99% of CEOs stated that sustainability will be crucial for their company success, since climate change is nowadays one of the most relevant factors in executive decision making. On the demand side, almost ¾ of customers globally claim are aimed at modifying their purchasing practices to be more environmentally friendly. In addition, a little less than half (45%) of investors actively take these considerations into account when making investing decisions⁶⁶.

The main positive impacts related to the adoption of Edge to Cloud systems from an environmental point of view will be the following:

- better work processes enabled by Edge to Cloud systems (and in general, by new levels of automation) will lead to more optimised usage of input materials and lower environmental impact. The introduction of a new cutting-edge technology will improve the whole system making it significantly more efficient;
- cloud is more environmentally friendly as unused processing power is usually diverted to other instances, whereas on-premises is exclusively utilised exclusively for internal processes;

⁶⁴ [Engie, Optimising data center energy consumption for “greener” digital technology, 2022](#)

⁶⁵ [McKinsey & Company, McKinsey Technology Trends Outlook 2022, 2022](#)

⁶⁶ [European Telecommunications Network Operators' Association \(ETNO\), Connectivity & Beyond, How Telcos Can Accelerate a Digital Future for All, 2021](#)

- the significant limitation of physical records (paper, printer ink toners, and so on) and the use of hardware resources on site (servers, server rooms, climate control, and so on) will help reducing the Carbon Footprint in the IoT domain;
- the ability for workers, encouraged and facilitated by the solution, to remote working will reduce the need to commute, and thus pollution;

the increasing relevance of Green Edge Processing as a market trend, meaning that, in the near future, more and more Edge-Cloud Processing solutions will be deployed focusing on sustainability and cost reduction, being them connected to renewable energy sources

4.3. Competitive Landscape

4.3.1. Relevant similar projects

Although on the current market there are no ready-to-market solutions offering the same service package as aerOS, however, as it has been shown in the first part of section 3.3.4., some European projects with similar goals, funded under the **Horizon Europe programme (Cluster 4, Destination 3: “Future European Platforms for the Edge: Meta-Operating Systems”)**, are being developed. In particular, a quick comparison must be done between aerOS and three interesting projects, that is to say: **ICOS (IoT to Cloud Operating System)**, **FLUIDOS (Flexible, scaLable secUre and decentralIseD Operating System)** and **NEMO (Data processing and communication platform)**. Nevertheless, it must be considered the fact that, as the following tables will show, aerOS will be one of a kind, since, despite similarities, it will be validated and tested in all the most significant market sectors, as opposed to competitors, that will only be validated in some of them: in fact, five pilots are already in place for aerOS testing (Manufacturing & Production, Renewable Energy Sources, Smart Buildings, Port Continuum, Machinery for Agriculture, Forestry & Construction), and, in the second phase of the project development, at least two more pilots, meaning the Automotive, Transportation & Mobility vertical and the health vertical will be dealt with.

From the point of view of the services actually offered, the ICOS project could be considered as the most complete and closest to aerOS: in fact, the intended solution includes device heterogeneity, continuum virtualisation, service orchestration, meta-Operating System, and Artificial Intelligence. Like aerOS, it is planned to be tested on the following fields: Agriculture, Energy, Automotive, Transportation & Mobility; yet, unlike aerOS, it will not face fields like Logistics, Industry 4.0, Smart Cities and Health.

Moving to the FLUIDOS project, the intended solution will be similar to aerOS in providing service orchestration and meta-Operating System, but it will not provide device heterogeneity, continuum virtualisation and Artificial Intelligence. Like aerOS, it is intended to be tested on the following fields: Agriculture, Energy and Logistics. Unlike aerOS, it will not deal with: Automotive, Transportation & Mobility, Industry 4.0, Smart Cities and Health.

Regarding the NEMO, that will develop the first integrated sensing data platform for noise and exhaust emission measurements for individual vehicles, the intended solution aims at providing continuum virtualisation, service orchestration, meta-Operating System and Artificial Intelligence, like aerOS does. Yet, unlike aerOS, it does not include device heterogeneity. From the point of view of the targeted vertical pillars, it will share with aerOS the fact that it will be tested on the following fields: Agriculture, Energy, Automotive, Transportation & Mobility, Industry 4.0 and Smart Cities. Unlike aerOS, it will not be tested on Logistics and Health.

	Device heterogeneity	Continuum virtualisation	Service orchestration	Meta - OS	AI
ICOS	✓	✓	✓	✓	✓
FLUIDOS			✓	✓	
NEMO		✓	✓	✓	✓
aerOS	✓	✓	✓	✓	✓

Figure 85: Services Comparison between aerOS and the HE projects ICOS, FLUIDOS and NEMO. Source: Own elaboration

	Agriculture	Energy	Logistics	Automotive, Transport & Mobility	Industry 4.0	Smart cities	Health
ICOS	✓	✓		✓			
FLUIDOS	✓	✓	✓				
NEMO	✓	✓		✓	✓	✓	
aerOS	✓	✓	✓	✓	✓	✓	✓

✓	Vertical pillars already targeted.
✓	Vertical pillars to be targeted in the near future through the open calls in the 2nd phase of aerOS.

Figure 86: Pilot Comparison between aerOS and the HE projects ICO. Source: Own elaboration

4.3.2. Business solutions

Although one of the major current trends on the market is the "all-in-1" formalisation of technological solutions, however it cannot be ignored the fact that some end users only need one or more, but not all the services offered by all-in-1, solutions such as aerOS: therefore, for reasons of habit towards the use of already existing solutions, as well as for economic reasons, aerOS will face the competition of some players regarding each of the individual services it offers; furthermore, these alternative solutions already on the market, combined and crossed, turn out to provide the same services, albeit fragmented into more than one solution and therefore less practical. The following paragraphs will list the main competitor providers in the respective sectors relating to the various services included in the aerOS solution.

4.3.2.1. Cloud Computing

Cloud computing services are provided by companies on a range that goes from full application development platforms to servers, storage, and virtual desktops. The main players with regards to cloud computing services offer the following solutions:

- **Amazon Web Services (AWS):** as a cloud web hosting platform, AWS provides fast, flexible, reliable and cost-effective solutions and offers a service in the form of building block which can be used to create and deploy any kind of application in the cloud. It is the most popular solution as it was the first to enter the cloud computing market.
- **Microsoft Azure:** launched in February 2010, this cloud platform by Microsoft is open-source, flexible, scalable and cost-effective, providing efficient services for development, data storage, service management and hosting solutions.
- **Google Cloud Platform:** part of the Google Cloud set of solution and products together with the G suite, it is one of the top providers in the field as it solves issue with accessible Artificial Intelligence and data analytics.
- **Oracle Cloud:** providing innovative and integrated cloud services, it represents one of the best cloud services providers that help to build, deploy, and manage workloads in the cloud or on premises. This solution is also regarded as efficient in helping companies to transform their business and reduce complexity. It uses modern technologies such as Artificial intelligence, chatbots and Machine Learning, and offers the next-generation mission-critical data management in the cloud.
- **IBM Cloud:** open and built with a strong suite of advanced and AI tools, it is regarded as one of the best cloud providers, spanning through public, private and hybrid environments. It proposes to the customers infrastructure as a service (IaaS), software as a service (SaaS) and platform as a service (PaaS).
- **Kamatera Performance Cloud:** very much similar to a physical server, the cloud server tool developed by Kamatera is considered flexible and cost-effective, since it operates in a virtual infrastructure cloud. It has 13 data centres across several countries: Canada, USA, Germany, United Kingdom, Israel and Hong Kong.
- **DigitalOcean:** a scalable computer service, this cloud platform offers add-on storage, security, and monitoring capabilities to run production applications in a quite easy way: it allows to deploy the user custom image, one-click app, or standard distribution.
- **Hostiger:** considered cheap and reliable and with a user-friendly interface, it offers cloud services that go from DDoS Protection and Cloud Firewall to Off-Site Backups.

Among the other notable players, the following should be mentioned: CloudSigma, LimeStone, Vultr, LiquidWeb, Linode, OVHcloud, Cloudways, and ScalaHosting.

4.3.2.2. Edge Computing

Edge computing is crucial for companies which wish to offer to their customers a more efficient modality to process and transmit data, solving two main problems: the instance for more IT infrastructure, and the significant number of unused data generated by edge points. The main players with regards to edge computing services offer the following solutions:

- **Amazon Web Services (AWS):** its cloud services are hybridised with edge in a model that includes IoT, AI, ML, robotics, analytics, and compute and storage capabilities. AWS also sells edge products, of which some of the best know are Alexa and Echo devices. AWS edge computing is good both for the industrial market and the commercial one.
- **Google Cloud Platform:** it offers a line of connected home products for edge computing and it also provides cloud computing services for managing edge data, mostly through its Cloud IoT Core service. Google uses Edge TPU hardware to run analytics and AI at the edge, and provides several other AI cloud services complementing its edge computing products.

- **IBM Edge Computing Platform:** it is based on OpenShift technology, and the Watson IoT applies its AI technology there.
- **ClearBlade:** released in 2020, the Edge Native Intelligent Asset Application by ClearBlade allows edge maintainers to connect IoT devices, define asset types, and build alert systems without needing any coding ability. This solution has been proved as efficient in sectors like: mining, facilities, oil and gas, rail, logistics, healthcare, and energy, but also in the public sector.
- **Dell EMC:** it provides edge-computing management and orchestration capabilities through OpenManage Mobile. Its hardware includes: the Mobile Edge portfolio (with cloud-enabled hardware for mobile or remote locations), the Enterprise Edge portfolio (consisting of the VEP460 Open uCPE platform), and the IoT Edge portfolio, with Edge Gateways for manufacturers, retailers, and digital cities.
- **EdgeConneX:** enabling tailored scalability and better network and IT connectivity, it offers EdgeOS, a self-service management application meant for high observability, with a single universal dashboard to manage it. The far edge use cases of this solution include artificial intelligence, AR/VR, IoT, low latency media streaming, connected and automated vehicles, immersive gaming, and machine learning.

Among the other notable players, the following should be mentioned: ADLink Technology, Intel, Mutable, HPE and Section.

4.3.2.3. Internet of Things (IoT) Software

The main players with regards to edge computing services offer the following solutions:

- **Arduino IDE IoT Software:** it is a cross platform integrated development editor written in Java that supports native microcontroller development using the C and C++ embedded programming languages.
- **Windows IoT:** Known previously as Windows Embedded, it is a popular Operating System for embedded systems that allows the development and the maintenance of IoT devices. Designed to boost the UWP app experience while providing a more accessible platform to develop such IoT software, it gives developers access to a vast and already established Windows ecosystem.
- **Android Things:** this IoT software offers a cutting-edge platform for IoT systems that require a considerably low memory footprint while still supporting numerous ARM-based architectures.
- **Microsoft Azure:** as mentioned before, this cloud computing platform allows to build, deploy, and test IoT software only on the cloud. It is used in the industry as a platform (PaaS) or infrastructure as a service (IaaS). The existence of the twin Azure Sphere, always by Microsoft but based on the famous Linux kernel, has to be mentioned.
- **MindSphere:** developed by Siemens, this IoT software enables everyday IoT devices to effectively collect and utilise cloud data in order to a better decision-making.
- **ROS (Robot Operating System):** it possesses a set of software libraries and tools for managing even the most complex robotic projects. It focuses on a modular developing paradigm and employs powerful abstractions to achieve its goal.

Among the other notable players, the following should be mentioned: Raspbian, MQTT, Thingspeak, Predix, OpenRemote, and Ending Thoughts.

4.3.2.4. Artificial Intelligence

Artificial Intelligence is currently at the core of the main part of actually significant solutions. So, as a general overview, the most relevant leaders in the Artificial Intelligence Market are listed below:

- **Google:** AI is profoundly integrated in almost the totality of products by Google: in fact, from smartphone assistants to image recognition and translation, a myriad of AI functionality hides within google apps that almost everybody uses daily. Not only Google search engine is powered by AI, but also its Ads and DoubleClick (both incorporate Smart Bidding, a machine learning powered automated bidding system), not to mention Google Maps (with its Driving Mode), YouTube (the Safe Content

uses ML techniques), Google Photo, Google Drive and Google Translate, which, for example, uses an artificial neural network called Google Neural Machine Translation (GNMT) in order to increase fluency and accuracy of translations.

- **Alphabet:** beyond Google, the company AI strategy includes fully owned ventures like Waymo and DeepMind. Alphabet is currently investing significantly in the development of AI solutions, with purposes that go from pharmaceutical and drug scouting to military support use.
- **Apple:** incorporating Machine Learning and AI into Apple products is aimed at improving the user experience: throughout iOS, macOS, iPadOS and watchOS there are several features and updates that have AI and ML at their heart (e.g., FaceID, Hand washing, Handwriting recognition).

Among the other players, it is important to mention: Albert Technologies, Amazon, Baidu, IBM, IPsoft, Microsoft Corporation, MicroStrategy, NVIDIA, Salesforce, Sentient Technologies Holdings, Qlik Technologies and Verint Systems (Next IT).

4.4. aerOS market position

To date, the challenge of seamlessly integrating various edge technologies into a homogeneous “continuum” remains unmet, since current centralised deployments store and process long-term data, relying on the cloud, but lack capabilities needed to handle cloud-centricity (more and more capable, cheap, small devices forming IoT ecosystems), latency, cloud costs (meaning price for storing and processing data which increases with resource use), network congestion for ever increasing world Internet traffic, smart devices, lack of security, privacy and trust.

All these elements are bringing cloud market and data economy to a turning point: while traditional cloud services move towards commoditisation, an innovation shift is required towards an IoT edge-cloud continuum, in which computing and storage resources can be located anywhere in the network, defining an expanded network compute fabric that spans over any fragment of the entire path between constrained devices and cloud (or clouds).

The specific type of Edge Computing used in aerOS is yet a unique solution in Europe, since it is not currently being used anywhere, at least not in any already ready-to-market competitor solution: aerOS is a proper and complete open-source interoperable cloud edge continuum solution, that deals with the whole ecosystem and not only with specific challenges.

More in detail, the breakthrough of aerOS lies in the fact that its architecture converges advances from:

- System view, since resources available in the compute continuum are geo-distributed and migrate over time (e.g., roaming smart devices or far edge nodes); compared to clouds, such resources are more heterogeneous and dynamically shifting.
- Data view, meaning structure and content of available (meta)data, including raw unstructured information, framed in standard, unified models and ontologies, compliant with prevalent communication interfaces to support data autonomy. That implies the ability of aerOS to manage data generated by heterogeneous sources and process them while taking into account instances for security, governance, provenance and traceability.
- Usage view: aerOS intends to orchestrate services in a more intelligent and automatic manner, advancing compatibility, portability, automation and interoperability of data within the IoT edge-cloud continuum, in order to face customer expectations and needs.

The vision behind aerOS includes goals such as:

- delivering common virtualised services to enable orchestration, virtual communication (network-related programmable functions), and efficient support for frugal, explainable AI and creation of distributed data-driven applications;
- exposing an API to be available anywhere and anytime (location-time independent), flexible, resilient and platform agnostic;
- including a set of infrastructural services and features addressing cybersecurity, trustworthiness and manageability. At the same time aerOS aims at: using context-awareness to distribute software task

(application) execution requests; supporting intelligence as close to the events as possible; helping execution of services using “abstract resources” (e.g., virtual machines, containers) connected through a smart network infrastructure; allocating and orchestrating abstract resources, responsible for executing service chain; providing support for scalable data autonomy.

Furthermore, aerOS sees as a purpose the perspective to leverage European leadership in automation systems in industry (where edge resides), proving how the whole field could actually benefit from decentralised, platform-agnostic IoT edge-cloud continuum data-processing ecosystem, while building competitive advantages e.g., reduced time to decisions; cost and time efficient, secure, trustworthy data sharing and control; semi-autonomous action taking; agile operations; sustainable, human-centric data processing, governance, and interoperability; reduced external traffic; and improved latency.

It must be underlined how edge components like micro data centres in the existing solutions are commonly used for disaster recovery, thus wasting a lot of potential. The aerOS approach has been set in order to be directly applicable to any vertical, cross-vertical business process, and several different physical or virtual platforms: unlike competitor solutions (for any of the services provided by aerOS), which are often too generic and not specifically targeted, aerOS is designed keeping in mind the fact that it will be tested in several use case scenarios, covering all the main verticals of the market (to the already present use case partners, more are about to join thanks to the open calls, e.g., in the field of Automotive, Transport and Mobility, and in Health sector), thus becoming a customised solution for several different sectors. It will answer the urgent need for a trustworthy, decentralised, autonomous, orchestrated solution, enabling bottom-up formation of compute continuum ecosystems, where hyper-distributed applications will be efficiently executed, within any selected “fragment” of heterogeneous physical infrastructure. It must be noticed that aerOS does not only possess a multiplicity of use case scenarios, but it was designed from the beginning of its conceptualisation through the combination of different perspectives, approaches and needs, incorporating visions that go from the one of telecommunication operators, to the one properly Cloud-based SMEs, of Edge hardware providers, of Academia, and more.

In perspective of governing the IoT edge-cloud continuum, aerOS integrates relevant technologies, elements of connectivity, IoT, AI, data autonomy and cybersecurity: the proposed meta operating system supports distribution and data sharing across the IoT edge-cloud continuum and enables orchestration of resources and services, by providing mechanisms for data processing and application of intelligence, also closer to where the data is produced.

Furthermore, from the point of view of IoT edge-cloud continuum orchestration, it is important to point out how aerOS delivers automated service orchestration, developing a robust high-performance algorithmic framework supporting full automation of service orchestration with adoption and fine-tuning of innovative AI/ML techniques (i.e. training time and accuracy), addressing different topologies, from hierarchical to fully distributed. aerOS will provide zero-touch orchestration leveraging ongoing standards and open-source initiatives, progressing shared learning between domains beyond the state of the art to speed up the learning process of AI/ML models.

Regarding smart networking, aerOS leverages its capabilities in the field (5G Native Exposed APIs NEF, SEL, CAPIF, programmable network fabric, etc.) in order to improve scalability, and real-time processing, within the network, by supporting data/knowledge distribution mechanisms, including automatic monitoring and dynamic (self-) configuration of the network, by means of SDN/NFV15 components to bring about smart network paradigm. Moreover, aerOS integrates Time Sensitive Networking (TSN) for timely delivery and reliability of critical control data in complex IoT edge-cloud continuum distributed infrastructures derived from the use cases.

On the crucial topic of containerisation and virtualisation, aerOS addresses the dynamic nature of IoT edge-cloud continuum constrained resources including re-configuration of smart networking elements, and re-evaluation of orchestrators, while developing effective mechanisms to distribute data across IoT edge-cloud continuum, so that the integrity and the performance of latency sensitive applications are not compromised.

Regarding the issue of dynamic data autonomy, aerOS is developed to comprehensively address to the matter through an integral data infrastructure, relying on current, well-established yet innovative solutions in use for IoT (CIM) and network telemetry (YANG), that would require extension and support for scaling, supporting user-defined policies integrated in data models, compositional models to define data processing topologies, for verification and validation, syntactic and semantic interoperability, runtime operation and management of data pipelines, and automated policy enforcement in heavily virtualised environments.

In perspective of Frugal Artificial Intelligence (FAI), aerOS contributes with the much-needed ability to explain in data pipelines within IoT edge-cloud continuum. Research was conducted on the topic of efficient implementation and orchestration of selected distributed frugal and/or explainable AI methods on resource constrained devices. Implemented AI modules are being validated in the laboratory and in actual use cases and complemented with lessons learned, assuring ease of use.

Concerning the concerns related to privacy in general, aerOS is intended to deliver the highest level of security, privacy and trust currently imagined, while keeping high performance, using lightweight SotA techniques, such as concise binary object representation signing and encryption, lightweight attestation, and lightweight consensus. Information, knowledge and decisions, shared amongst peers, will remain trustable thanks to traceability and accounting mechanisms, while leveraging a newly defined DevPrivSecOps methodology, including Security and Privacy in the DevOps processes. By design, aerOS supports, thanks to its modular architecture, any existing and future cybersecurity mechanism. In conclusion, although the aspects of data security and data storage are already covered by existing solutions, the necessity to implement solutions which are more elastic and flexible about data was still felt, thus leading to aerOS conceptualisation.

From the point of view of technological standardisation, aerOS is intended to become a European Standard, and this by itself constitutes a significant selling point on the market: starting from the fact that having aerOS as a public, reference standard will shape how future European technology will be developed, it must be pointed out how even the idea of possessing a technology standardised by UE makes the solution implementation a much safer, reliable and trustable decision. Furthermore, the adoption of a standardised EU technology comes with many benefits: for example, the awareness of the fact that other European companies use the same technology enables intercommunication and interoperability among companies or entities (e.g.: internet). The push to technology evolution is impossible to ignore: if a technology is standardised and vastly used in society it is likely to evolve into improved iterations (e.g.: Ethernet, Wi-Fi), for everyone's benefit, while a proprietary technology (contrary to a UE-standardised one) may not ever succeed in evolving, due to potential lack of users that discouraging companies.

From the social and environmental point of view, aerOS has been conceived in order to show a meticulous care towards the sustainability of resources, thus responding to an ever-increasing social instance that is not met by other existing solutions, usually connected to standard, and yet not renewable, energy sources.

4.5. Verticals addressed in aerOS - Market trends

4.5.1. Manufacturing – production

The market of smart manufacturing is constantly growing, and it is currently valued at \$ 88.7 billion (2021). According to forecasting, it is expected to be valued at \$ 228.2 billion by 2027, growing at a CAGR of 18.5% in the 2022-2027 time period⁶⁷.

The EU27 produces 22% of the world manufactured products, resulting in an annual trade surplus of € 421 billion. This result is also derived from the strong relevance of the Research and Development investments of the EU research institutions and enterprises in the sector that allows it to advance in the evolving smart manufacturing sector. Indeed, European research institutions and businesses, especially small and medium-

⁶⁷ [Bloomberg, Smart Manufacturing Market worth \\$228.2 billion by 2027 - Exclusive Report by MarketsandMarkets™, 2022](#)

sized, are important contributors in the field of Research and Innovation. Manufacturing enterprises account for 49% of innovation spending and 64% of private sector R&D spending in Europe⁶⁸.

The manufacturing sector plays a key role for the European Union economy; indeed, it accounts for the 83% of EU export⁶⁹ and represents a relevant part of the European Union economy, making up 8.9% of all enterprises and employing 30.2 million people, constituting the 22.9% of the total employment⁷⁰. The impact on the market is also crucial, as the sector generated almost € 2 billion of value-added amounting to the 29.2% of the total non-financial business economy EU value added⁷¹.

The manufacturing industry has experienced significant changes since the First Industrial Revolution, mutating along with the development of new technology. In fact, there was a steady transition from the water and steam power, that permitted mass manufacturing, to other power sources (e.g., oil, gas, and electric power) and an initial level of automation. After that, in the middle of the 20th century, developments in new technologies began, including computers, improved communications, and data analysis: it led to the start of a genuine process automation and to the gathering of data from processes. The Fourth Industrial Revolution, which is now being brought about by the discovery of new technologies, is centred on using intelligent machines and factories in addition to growing automation. This phenomenon promotes improved decision-making and gives access to a large amount of accurate data that supports more fruitful and efficient production of commodities throughout the value chain, also allowing companies to deploy an enhanced sustainability strategy improving circular economy. Increased flexibility enables producers to deploy mass customisation, in order to better satisfy consumer tastes and preferences⁷².

Currently, the sector shows a wage-adjusted labour productivity ratio⁷³, that measures the labour productivity, slightly above the EU average, with a value added per person employed in the EU manufacturing sector amounting to 150.2%, compared to an average for non-financial business economy of 142.6%⁷⁴. Despite this result, it is fundamental for manufacturing companies to keep invest in increasing digitisation of the sector as a lever to enhance its productivity.

In this regard, digitisation is becoming increasingly more important for companies to maintain and improve their competitiveness on the market: indeed, almost a quarter of CEOs state that advancing the digitisation and connectivity of all their companies' functional areas is their top priority over the next three years. In addition to this, sustainability issues are growing in importance, as 16% of CEOs think that to achieve their company-growth objectives it will be fundamental to integrate ESG reporting into their measurement and reporting processes⁷⁵.

Indeed, for the manufacturing sector it will be pivotal in the next few years to push forward their decarbonisation strategy, since the industrial sector in general (comprising both the manufacturing and the other industrial sectors) is showing the greatest variability among different countries decarbonisation readiness. Currently, the countries showing the highest rate of decarbonisation readiness⁷⁶ of the industrial sector are Japan (82%), Norway (72%), UK (71%), Germany and Denmark (both 70%)⁷⁷. In this regard, as cutting-edge digital

⁶⁸ [European Commission, Advanced manufacturing, 2021](#)

⁶⁹ [Eurostat, Extra-EU trade in manufactured goods, 2022](#)

⁷⁰ Taking into consideration NACE sections within the EU's non-financial business economy

⁷¹ [Eurostat, Manufacturing statistics – NACE Rev.2, 2022](#)

⁷² [IBM, What is Industry 4.0?](#)

⁷³ It is defined as value added divided by personnel costs which is subsequently adjusted by the share of paid employees in the total number of persons employed, or more simply, apparent labour productivity divided by average personnel costs (expressed as a ratio in percentage terms) (Definition by [Eurostat, Glossary: wage-adjusted labour productivity ratio](#))

⁷⁴ [Eurostat, Manufacturing statistics – NACE Rev.2, 2022](#)

⁷⁵ [KPMG, Global Manufacturing Prospects 2022 The CEO view: Supply chain resiliency helps achieve a twin transformation, 2022](#)

⁷⁶ The decarbonisation readiness Index is a tool that compares the progress of 32 countries in reducing the greenhouse gas emissions that cause climate change and assesses their preparedness and ability to achieve Net Zero emissions of these gases by 2050 (Definition by: [KPMG, Net Zero Readiness Index 2021, 2021](#))

⁷⁷ Ibidem

technology support sustainability activities, the Industry 4.0 (I4.0) concept opens the path for the circular economy, since digital technologies are crucial for facilitating the shift to that goal.

4.5.1.1. Current trends and problems

For a long time, the industrial sector has stressed the need to boost efficiency while simultaneously ensuring that its operations are resilient and agile, and this emphasis has only increased in the aftermath of the Covid-19 epidemic. Indeed, over the past several years, manufacturers have expanded their digital investment and hastened the adoption of innovative technology. Companies with higher digital maturity have demonstrated stronger resilience: in this regard, we can see how business digital transformations are accelerating and how IoT and data analytics are becoming more and more important.

The increasing application of digital technologies in the manufacturing production, such as advanced sensors, embedded software and robotics, allows smart factories to collect and analyse data in order to improve their productivity and to enhance their decision-making processes. The graph below shows which are the technologies on which manufacturers will focus the most in the coming year in order to increase the operational efficiencies of their factories.

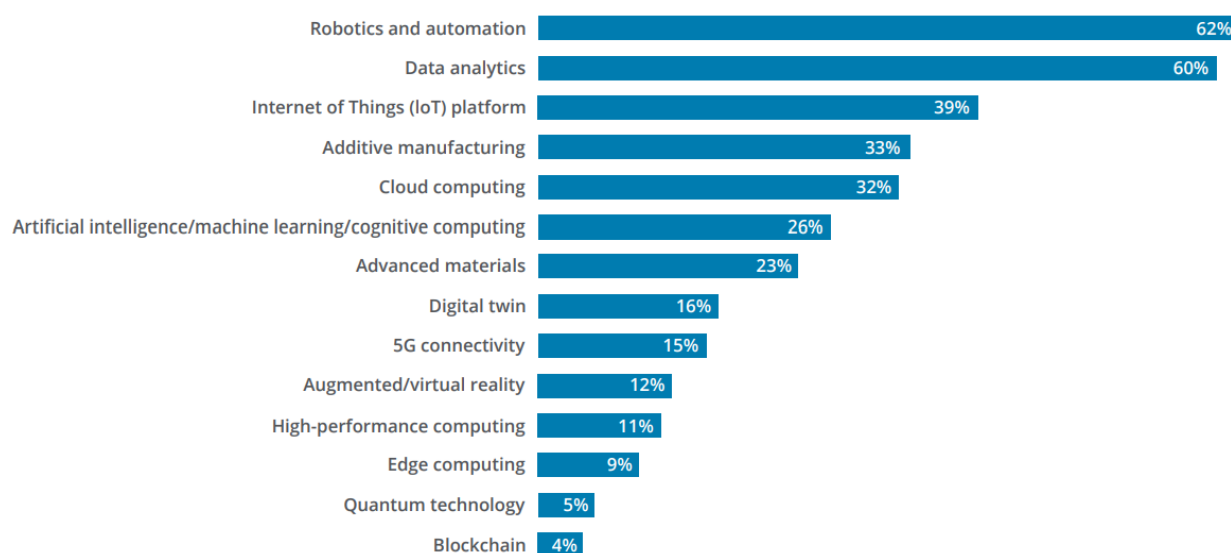


Figure 87. Surveyed manufacturers plan to focus on a range of technologies to increase operational efficiencies over next 12 months. Source: Deloitte, 2023 Deloitte manufacturing outlook survey, 2022

Organisations will place a high priority on IoT edge computing continuum in order to increase productivity and the efficacy of data analysis. In this approach, manufacturers will be able to increment their ability to lower emissions, build more resilient supply chains, and boost production outputs while quickly identifying issues and preventing downtime.

Developing an IoT edge-cloud continuum can help manufacturers in creating a fully automated environment and in providing real time, high bandwidth, and low latency connectivity in temporary networks for the execution of advanced automation application applied to swarm machines. Predictive analytics, automation of control and monitoring processes, increased production, elimination of low latency, and logistics optimisation are just a few of the network functions that encourage manufacturers to use an infrastructure with the bandwidth to handle the enormous amounts of data that endpoint devices send and receive. Through enhanced robotics and machine-to-machine communication closer to the source, edge computing enables manufacturers to integrate automation in factories and supply chain activities. This improves low latency and leads to faster analysis and correction. All of these features not only converge in a higher product quality, but they also lead to a higher productivity, through an increased yield and reduction of wastes and costs decrease.

Manufacturing production can get substantial benefits from the advances in data analytics as well as other emerging technologies such as Artificial Intelligence and Machine Learning: indeed, if implemented, these solutions can provide high values to the companies. McKinsey & Company⁷⁸ has estimated that the application of the emerging technologies related to the Industry 4.0 paradigm in manufacturing factories can bring high returns in different area of companies, amounting to:

- 15-20% inventory-holding cost reduction;
- 15-30% labour productivity increase;
- 30-50% machine downtime reduction;
- 10-30% throughput increase;
- 85% forecasting accuracy improvement;
- 10-20% cost-of-quality improvement.

In this sense, the implementation of a meta operating system for the IoT edge-cloud continuum is of fundamental importance to get the full potential from these technologies. Indeed, it is estimated that the application of IoT solution in manufacturing factories can have an economic impact between \$ 1 and \$ 2.3 trillion by 2030: given its characteristics, the manufacturing sector is particularly well suited to the applications of IoT technologies and is actually anticipated to be the sector that will benefit the most from their application by capturing the greatest economic value, accounting for 26% of the total economic value from the application of IoT technologies in 2030⁷⁹. The table below shows the use cases for the manufacturing sector that will provide the highest estimated economic value.

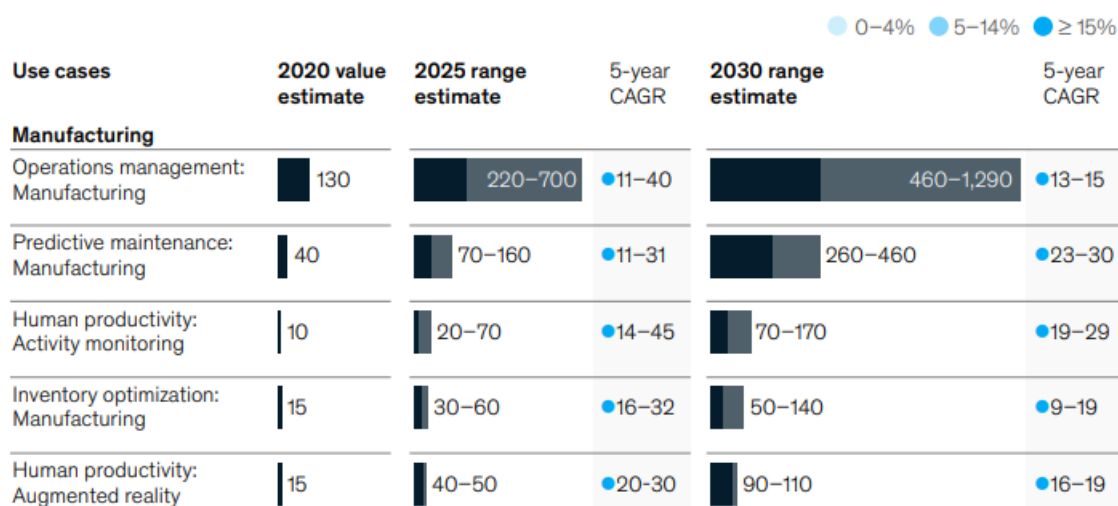


Figure 88. Estimated economic value by use case, 2020–30, \$ billions. Source: McKinsey, *The Internet of Things: Catching up to an accelerating opportunity*, 2021

Although IoT and other emerging technologies applications (e.g., Artificial Intelligence, Machine Learning) are expanding rapidly in the manufacturing sector, and even though the underlying technologies are sufficiently advanced and executives are aware of the enormous benefits of using these technologies with the potential game-changing impact they unleash, it has been documented that many of these cutting-edge applications never progress past pilot stages. In reality, expanding innovative technologies is a challenge for many businesses. McKinsey has stated that, due to several concerns with interoperability and cybersecurity, 70% of industrial businesses have been unable to grow their pilot programs⁸⁰.

⁷⁸ McKinsey & Company, *Capturing the true value of Industry 4.0*, 2022

⁷⁹ McKinsey & Company, *The Internet of Things: Catching up to an accelerating opportunity*, 2021

⁸⁰ Ibidem

4.5.2. Renewable energy sources

The energy and utilities market are projected to be the fastest-growing market in the 2022-2032 period, showing a CAGR of 29%. The main driver of this growth is the increasing importance of energy grids, in particular, with reference to remote monitoring and controlling technologies aimed at energy efficiency, which is becoming increasingly important in the market⁸¹.

The EU is becoming increasingly committed to sustainable development. As a result, with the long-term objective of safeguarding the environment, a rising number of directives and strategies, including Europe 2020⁸², the 2030 climate and energy framework⁸³, the 2050 long-term plan⁸⁴, and the European Green Deal⁸⁵, that has been translated into law by the European Climate Law⁸⁶, have been developed in recent years. According to the European Climate Law, the EU must reduce its greenhouse gas emissions by at least 55% by 2030 compared to 1990 and achieve climate neutrality by 2050. In order to achieve these objectives EU member states must implement specific steps to cut emissions and decarbonise the economy. The “Fit for 55” package⁸⁷ aims at aligning EU legislation to the proposed climate transition, reducing net greenhouse gas emissions and achieving climate neutrality.

The energy sector has a fundamental role in this sense, as over 75% of the greenhouse gas emissions in the EU come from the energy sector. Thus, to reach the EU decarbonisation targets, increasing energy from renewable energy sources will be crucial in both reducing GHG emissions by at least 55% by 2030 compared to 1990 and in making the EU a climate-neutral continent by 2050. Indeed, the benefits coming from the usage of renewable energy are generally acknowledged and go from the reduction of greenhouse gas emissions and the diversification of energy supplies to a significantly reduced dependency on fossil fuel like oil and gas, and even to the creation of new job positions for the implementation of a still partially unexplored field like the ‘green’ technology one.

As of 2020, 22.1% of the energy consumed in the EU came from renewable resources almost 2 points above its target and with a significant increase in comparison with the 9.6 % of 2004. In order to achieve the EU decarbonisation targets, the EU set a new objective for 2030, rising the target for the total energy consumption from 32% to 40%, requiring Member States to almost double the current share of renewable energy consumption⁸⁸. Eurostat investigation published in January 2022 presented the market for renewable energy sources as described below⁸⁹.

In the 2010-2020 decade, the increased amount of electricity coming from “green” sources mainly regarded wind power (33% of the total), solar power (14%) and solid biofuels, including renewable wastes (8%). In 2020, renewable energy sources made up 37.5% of gross electricity consumption in the EU, up from 34.1% in 2019. The growth in electricity from solar power has been impressively significant, rising from 7.4 TWh in 2008 to 144.2 TWh in 2020.

⁸¹ [Bloomberg, Burgeoning Adoption of Internet of Things \(IoT\) to Steer Global Edge Computing Market Past US\\$ 69 Bn through 2032, 2022](#)

⁸² [European Commission, Communication from the Commission, Europe 2020, A strategy for smart, sustainable and inclusive growth, 2020](#)

⁸³ [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions a policy framework for climate and energy in the period from 2020 to 2030](#)

⁸⁴ [Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, the Committee of the Regions and the European Investment Bank a Clean Planet for all a European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy](#)

⁸⁵ [European Commission, A European Green Deal](#)

⁸⁶ [Regulation \(EU\) 2021/1119 of the European Parliament and of the Council of 30 June 2021 establishing the framework for achieving climate neutrality and amending Regulations \(EC\) No 401/2009 and \(EU\) 2018/1999 \(‘European Climate Law’\)](#)

⁸⁷ [Council of the EU and the European Council, Fit for 55](#)

⁸⁸ [European Commission, Renewable energy targets](#)

⁸⁹ [Renewable energy statistics - Statistics Explained \(europa.eu\)](#)

Among the EU Member States, as of 2020, there is a high heterogeneity, with countries where electricity consumed generated from renewable sources accounts for more than 70%, such as Austria (78.2%) and Sweden (74.5%); yet, a crucial role was also played by Denmark (65.3%), Portugal (58%) and Latvia (53.4%). In addition, Norway and Iceland produced more electricity from renewable sources than they consumed in 2020, leading to a share higher than 100%. On the opposite, there are other countries such as Malta (9.5%), Hungary (11.9%), Cyprus (12.0%), Luxembourg (13.9%) and Czechia (14.8%) that show very low percentages.

Figure 89:

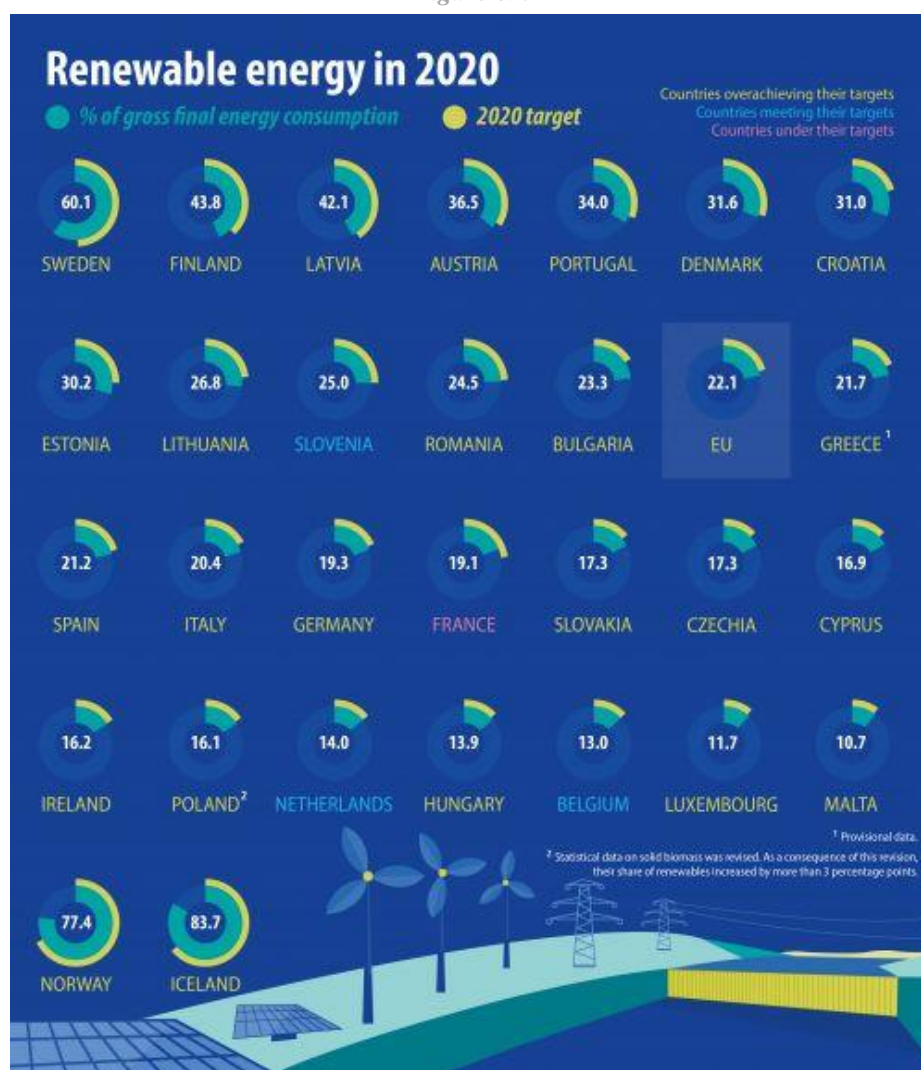


Figure 90. Share of energy from renewable sources, 2020 (% of gross final energy consumption) Source: Eurostat, Renewable energy statistics, 2022

4.5.2.1. Current trends and problems

Electricity is a crucial component of contemporary technology, and its importance will grow over the coming years. In fact, it is predicted that by 2050 global power consumption would rise by about 70 percent, rising from 25 to 42 terawatt-hours. The energy mix will vary with a substantial growth in the usage of renewable energy, which will account for 56% of all power produced by 2050. Power plants are connected and forms grids to transmit electricity to cities, homes and businesses. In order to make these systems more efficient, safe and reliable, over time several grid management applications have been developed. Among them the most time-responsive are the ones that employ cloud and edge computing, IoT and AI technologies. The increasing privacy,

security and reliability requirements, coupled with an increasing demand for energy consumption and the related need to avoid inefficiencies in the systems, requires the adoption of innovative technologies⁹⁰.

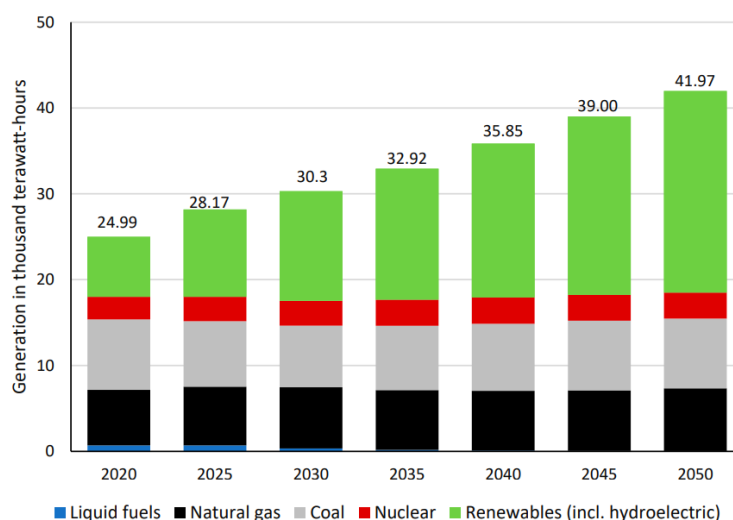


Figure 91. The forecast of worldwide electricity consumption, for the period of 2020 to 2050. Source: Minh, Q.N.; Nguyen, V.-H.; Quy, V.K.; Ngoc, L.A.; Chehri, A.; Jeon, G. *Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. Energies* 2022

To match the increasing electricity demand and the wide-scale adoption of renewable resources energy, it will be important to further the digitalisation of Renewable Energy systems in order to enhance automatization of decision making and providing innovative information. Indeed, there are some barriers to the wider adoption of Renewable Energy sources. The main obstacle lies in the unpredictable climatic circumstances that may produce intermittent energy generation. In addition, the decentralised and distributed configuration of Renewable Energy structure in the local area (such as small-scale facilities such as solar rooftop installations, biomass generation, wind farms, or small hydropower plants) creates management and connection challenges across networks, which complicate information flows in the new energy architecture and make it challenging to monitoring the grids at large scale for the community or city-wide adoption⁹¹.

To the aim of renewable energy sources, the implementation of a meta operating system for the IoT edge-cloud continuum would make it possible to get all the benefits of the cloud computing power while avoiding its drawbacks and get full advantage from the data obtained from the use of leak sensors, temperature sensors, vibration sensors, humidity sensors, video sensors. Indeed, it would be possible to get the data from the IoT devices and sensors and compute and process them at the edge instead of at the cloud as it is traditionally done. It will allow the decision-making process to be performed at the edge, accelerating it. Additional benefits would include quicker service response times, low transmission latency, the ability to make decisions instantly, and a reduction in bandwidth traffic burden. After the data are aggregated at the Edge Computing layer, only the key information would be sent to the Cloud Computing layer for computing, statistics, and storage. This architecture reduces the distance between databases and end users, bringing cloud capabilities closer to the customers. It also addresses the issue of governments and electrical companies sending their data to foreign data centers due to privacy concerns⁹².

Renewable Energy sources sector and the application of IoT may also benefit from the of other emerging technologies such as Artificial Intelligence and Machine Learning. More specifically the edge AI, defined as “the use of AI techniques embedded in Internet of Things (IoT) endpoints, gateways and other devices

⁹⁰ Minh, Q.N.; Nguyen, V.-H.; Quy, V.K.; Ngoc, L.A.; Chehri, A.; Jeon, G. *Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. Energies* 2022, 15, 6140. <https://doi.org/10.3390/en15176140>

⁹¹ Rahul Mishra, B. Koteswara Rao Naik, Rakesh D. Raut, Mukesh Kumar, *Internet of Things (IoT) adoption challenges in renewable energy: A case study from a developing economy*, *Journal of Cleaner Production*, Volume 371, 2022, 133595. <https://doi.org/10.1016/j.jclepro.2022.133595>.

⁹² Ibidem

computing data at the point of use”⁹³. It can help in implementing intelligent forecasting that will play an increasingly relevant role in the proper development of the Renewable Energy sector. Indeed, it can ensure a better energy resource generation, distribution, and management by combining historical data, weather patterns, grid health, and other information through complex simulations. Other major applications of Artificial Intelligence in the Renewable Energy sector concern grid management, meaning the ability to predict energy consumption in households through Artificial Intelligence analysis of several data such as the specific period of the year, maintenance needs prediction⁹⁴.

Despite the undeniable potential of Artificial Intelligence application in the renewable energy market, there exist some vulnerabilities. Indeed, the reliance on AI technologies could lead to cyber-attacks. These, together with other potential problems that could hinder the successful deployment of AI technologies in the sector relate to performance such as data bias, audit and ongoing verification of algorithms and to technology barriers such as a potential lack of reliable connectivity, a fundamental aspect for this kind of application, especially in rural and under-served areas. Finally other barriers concern a lack of trust from consumers and regulatory hurdles⁹⁵.

4.5.3. Port Continuum

The maritime commerce industry has grown consistently, over the past 20 years, at a rate of 2.9% yearly. This is also a result of the constant rise in trade globalisation and of the expansion of the manufacturing industry, both of which depend heavily on ports. In fact, ports show to be an important component of commerce, since they not only guarantee a high degree of trustworthiness, but also happen to be a cost and energy-effective method. To date, 80% of all commerce volume occurs through the sea, making up a sizable portion of global trade. Transport in containers is crucial in this context: it accounts for 60% of trade value and 35% of overall trade volume⁹⁶.

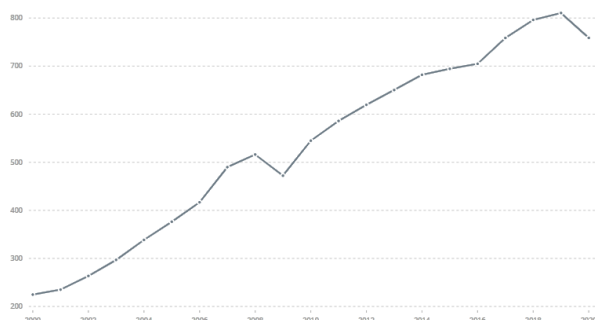


Figure 92: Worldwide container port traffic (2000-2020) – TEU (Twenty-foot Equivalent Unit)⁹⁷. Source: The World Bank, Container port traffic

The port environment particularly lends itself to automation, given its conformation coupled with the high level of repetitiveness of activities. Port automation is, indeed, a trend that started back in the 1990s in Europe and, as of today, around 53 container terminals in the world present a partial automated environment, accounting for around 4% of all container terminal capacity worldwide. Despite these data, to date, there are no fully-automated container ports in the world: furthermore, according to predictions, automation might cause around 90% of the dock job that exists now to vanish by 2040. Currently, there are discrepancies among the location of the ports that are most automated: in particular, the container yard, that is where the bulk of automated systems are

⁹³ [Gartner, Innovate With Edge AI, 2019](#)

⁹⁴ [Forbes, How Artificial Intelligence And Machine Learning Are Transforming The Future Of Renewable Energy, 2021](#)

⁹⁵ [EY, Why artificial intelligence is a game-changer for renewable energy, 2020](#)

⁹⁶ [The World Bank, The Container Port Performance Index 2021 : A Comparable Assessment of Container Port Performance, 2022](#)

⁹⁷ Data on the vertical axis are in millions.

located, is the most automated section of ports, while the transit between quay and yard is only automated in a small number of ports. Finally, quay cranes are not fully automated at any terminal⁹⁸.

Hence, smart and automated ports are increasingly important in the ports landscape. An automated port, as shown in the figure below, consists of the following main areas: the berthing area at the quayside, equipped with quay cranes for unloading and loading containers, the travelling area of Automated Guided Vehicles (AGVs), used by AGVs to move containers from the berthing area to the storage yard, the storage yard that stores import and export containers before further delivery by trucks or trains. An automated port is, then, referred to as a “smart port” when it employs big data, blockchain, the internet of things (IoT), and other cutting-edge technologies to boost efficiency and competitiveness⁹⁹.

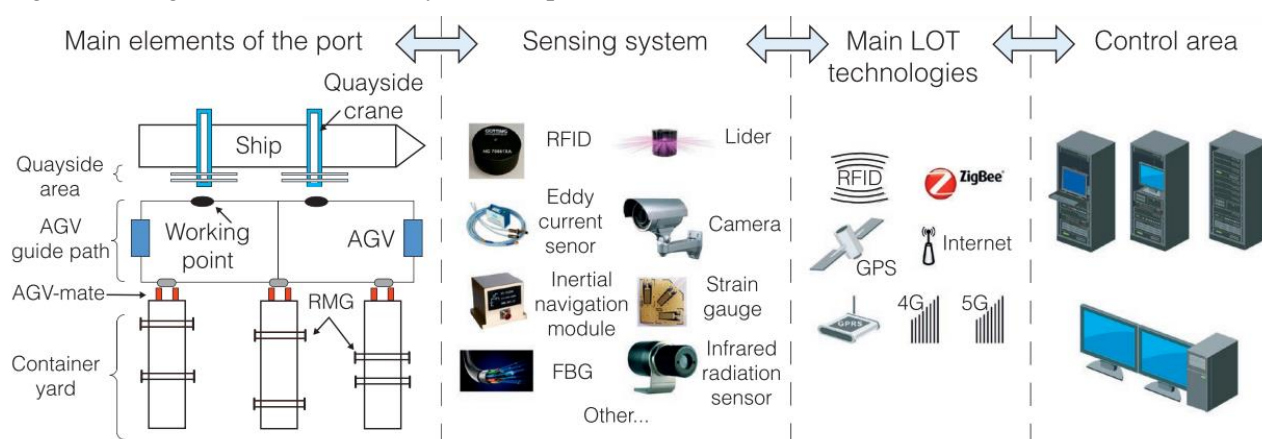


Figure 93: The layout of an automated container terminal. Source: Yang, Yongsheng & Zhong, Meisu & Yao, Haiqing & YU, Fang & Fu, Xiuwen & Postolache, Octavian. *Internet of things for smart ports: Technologies and challenges*. IEEE Instrumentation & Measurement Magazine, 2018

Thanks to their configuration, automated container ports are able to automate several processes in the terminals. The first processes that have been automated in ports are the decision-making processes that, without automating the port assets themselves, enhance managerial and performance aspects of the port. The automation of gates is another type of automation put in place for the container port terminal automation: it is implemented to streamline operations at gates, as these systems make it possible to automatically identify drivers and containers, process bills of lading, direct drivers to their designed spot for the loading and unloading operations, and implement sophisticated appointment systems. The use of modern technologies in ports allows also to automate the tracking and tracing of several components of the terminals (e.g., ships, cranes, containers and yard equipment), allowing an enhanced concertation of equipment and operations. Finally, also the loading and unloading operations of the containers can be automated: indeed, the automation of the equipment used in the container terminal to carry out these operations is crucial as the high volume of containers passing through the port requires a significant use of them. In particular, Quay Cranes (QCs), AGVs, and Yard Cranes (Ycs) are used for these operations¹⁰⁰.

4.5.3.1. Current trends and problems

Ports productivity is influenced by the shortage of space: this requires ports to commit at improving space productivity. Given the difficulties or even the impossibility for ports to expand in size, it will be necessary to improve productivity through increasing investments in innovation and automation, and this will result in a higher use of robotics and IoT.

Increased automation would lead to several advantages for ports, such as an increase in safety, a lower rate of human-related disruptions and higher predictability of performances; it would also enhance decision-making

⁹⁸ [ITF, “Container Port Automation: Impacts and Implications”, International Transport Forum Policy Papers, No. 96, OECD Publishing, Paris, 2021](#)

⁹⁹ [Yang, Yongsheng & Zhong, Meisu & Yao, Haiqing & YU, Fang & Fu, Xiuwen & Postolache, Octavian. Internet of things for smart ports: Technologies and challenges. IEEE Instrumentation & Measurement Magazine, 2018](#)

¹⁰⁰ [Notteboom, T., Pallis, A., & Rodrigue, J.-P. Port Economics, Management and Policy \(1st ed.\). Routledge, 2022. https://doi.org/10.4324/9780429318184](#)

capabilities. These features will reflect in practical advantages for ports, both in cost reduction, with a foreseen decrease in operating expenses by 25-55%, and in both safety and performance gains. For instance, it is predicted that enhanced automation would lead to a marked increase in productivity by 10-35%¹⁰¹.

The use of edge and cloud computing technologies will also enable the digital transformation of ports and the automation of crucial activities. In reality, this will make it possible to track and manage significant numbers of containers in real time optimising all of the port activities (e.g., management of docks, tug operations and pilotage), while also enhancing safety and incident prevention. 5G network will also play a fundamental role in the automation of ports as it can provide ultra-fast and low-latency communications, improving efficiency and productivity through the use of key systems such as drones and AGVs¹⁰². 5G will work as a key enabler for the deployment of the technologies needed to make ports smarter and to connect workers, machines, cranes etc. across the port. It will serve as a communication platform for edge and cloud computing, Artificial Intelligence and Machine Learning as well as for other technologies such as Digital Twins. Also, Artificial Intelligence will enhance smart ports efficiency: besides the role played in the introduction of robotics and Automated Guided Vehicles (AGVs), it has the potential to provide additional value to ports through advanced processing of past and real-time data coming from IoT. Finally, it will fuel the adoption of a pervasive smart port paradigm, helping to replace traditional equipment with the automated counterparts.

In a similar way to what was set out above, these sorts of technologies, even if they are cutting-edge and essential to maintaining the industry growth in the years to come, would expose ports to the potential of being shut down or having data stolen, leaving them vulnerable to cyber-attacks. Other obstacles to the adoption of automation at ports include the significant upfront costs associated with putting automation in place, as well as operational issues, such as skill gaps, bad data, and difficulties in handling exceptions¹⁰³. Other factors are also influencing the smart ports up taking: indeed, even though the undeniable necessity for ports to increase their automation and digitalisation with its predisposition, the sector has adopted automation solutions more slowly compared to other industries, such as mining. This phenomenon is due to a number of issues affecting ports: for instance, automation has not yet produced the economic benefits that were anticipated. In some automated ports, the return on invested capital of the assets is underperforming in the face of very high initial investment costs. Indeed, they show a return-on-investment rate up to 1 percentage point below the industry norm of about 8%, and operating expenses lowering of just 15%-35% against an expected decrease of 25%-55%, while productivity even shows a decline by 7 to 15% against an expected increase of 13-35%¹⁰⁴.

McKinsey¹⁰⁵ identifies the causes of these disappointing results in four main factors. The first relates to the shortage of skilled professionals required to carry out the tasks requested by the new automatised setting, also considering that training that would need a very long time to be undertaken correctly, with an esteem of around 5 years to be completed. The second factor concerns the fact that a proper automation of ports requires a good quality of data, structured data and data analytics, but these are currently lacking in ports, leading to inefficiencies. Another fundamental factor regards the presence of siloed operations among functions in ports, while for a comprehensive and efficient automation the integration and the collaboration across functions have a key role. Finally, another relevant aspect is the high number of exceptions characterising ports: in these cases, it is crucial to smooth processes before automating them, in order to avoid the presence of muddled operations¹⁰⁶. Therefore, with the aim of gaining access to an increase in productivity derived from the transition to the smart port paradigm, it will be necessary to fix all of these gaps.

¹⁰¹ [McKinsey & Company, The future of automated ports - The challenges are significant, but careful planning and implementation can surmount them, 2018](#)

¹⁰² [NearbyComputing, Ports & Container Terminals Use Case, 2021](#)

¹⁰³ [McKinsey & Company, The future of automated ports - The challenges are significant, but careful planning and implementation can surmount them, 2018](#)

¹⁰⁴ Ibidem

¹⁰⁵ Ibidem

¹⁰⁶ Ibidem

4.5.4. Smart Building

Buildings account for a high share of total electricity and final energy consumption, amounting to the 71% and 39% of the total consumption in urban areas. These also contributes to the high emissions caused by the building sector. Indeed, it has a high impact on the emissions of the urban areas, amounting to the 40% of the total CO₂ emissions happening there¹⁰⁷.

In Europe, buildings will also play a key role in the pathway to reach the aforementioned decarbonisation target, since buildings account for the 40% of total energy consumption. In order to reach the net-zero emissions target at the European Union level by 2050, it will be necessary a profound renovation of the building assets. Currently, the renovation rate of the existing building assets amounts to about 1% yearly, while to successfully reach the decarbonisation objectives it should rise to the 3%¹⁰⁸. In addition, new buildings in the European Union are prohibited from installing fossil fuel-based systems and 100% of building energy demand must be provided by renewable energy resources.

Smart buildings have been defined by the Buildings Performance Institute Europe (BPIE)¹⁰⁹ as “highly energy efficient and cover their very low energy demand to a large extent by on-site or district-system-driven renewable energy sources. A smart building (i) stabilises and drives a faster decarbonisation of the energy system through energy storage and demand-side flexibility; (ii) empowers its users and occupants with control over the energy flows; (iii) recognises and reacts to users and occupants needs in terms of comfort, health, indoor air quality, safety as well as operational requirements”.

The application of Artificial Intelligence and Internet of Things devices to building allows to propose a new paradigm for the application of human-machine interaction in general. This paradigm shift has the potential to bring several advantages to the building occupants and to enhance their experience, the building operational efficiency, and to optimise space and asset utilisation. In addition to provide real-time insights based on data collected, AI technologies can also offer useful prediction and respond to anomalies.

These advantages are felt also by the CIOs that, as detected by an IBM Institute for Business Value (IBV) study, stated for the 76% that automation in facilities and asset management could positively impact on operational efficiency, and for the 70% that data provided by intelligent machines may provide insights with the potential to enhance decision-making processes.

The advantage provided by smart building derives by the seamless integration of Internet of Things and Artificial Intelligence: indeed, AI technologies are capable of integrating and processing the huge amount of data coming from the IoT devices, in order to apply what they learn to the improvement of the energy performance of the buildings and to the optimisation of their general ecosystem (e.g., monitoring the park lot utilisation, lighting, maintenance needs). The image below summarises the capabilities of a smart building¹¹⁰.

¹⁰⁷ Farzaneh, H.; Malehmirchegini, L.; Bejan, A.; Afolabi, T.; Mulumba, A.; Daka, P.P. *Artificial Intelligence Evolution in Smart Buildings for Energy Efficiency*. *Appl. Sci.* 2021, 11, 763. <https://doi.org/10.3390/app11020763>

¹⁰⁸ J. Al Dakheel, C. Del Pero, N. Aste, F. Leonforte, *Smart buildings features and key performance indicators: A review*, *Sustainable Cities and Society*, Volume 61, 2020, 102328, ISSN 2210-6707. <https://doi.org/10.1016/j.scs.2020.102328>.

¹⁰⁹ *Buildings Performance Institute Europe (BPIE), opening the door to smart buildings, 2017*

¹¹⁰ *IBM Institute for Business Value, Building intelligence into buildings, 2018*

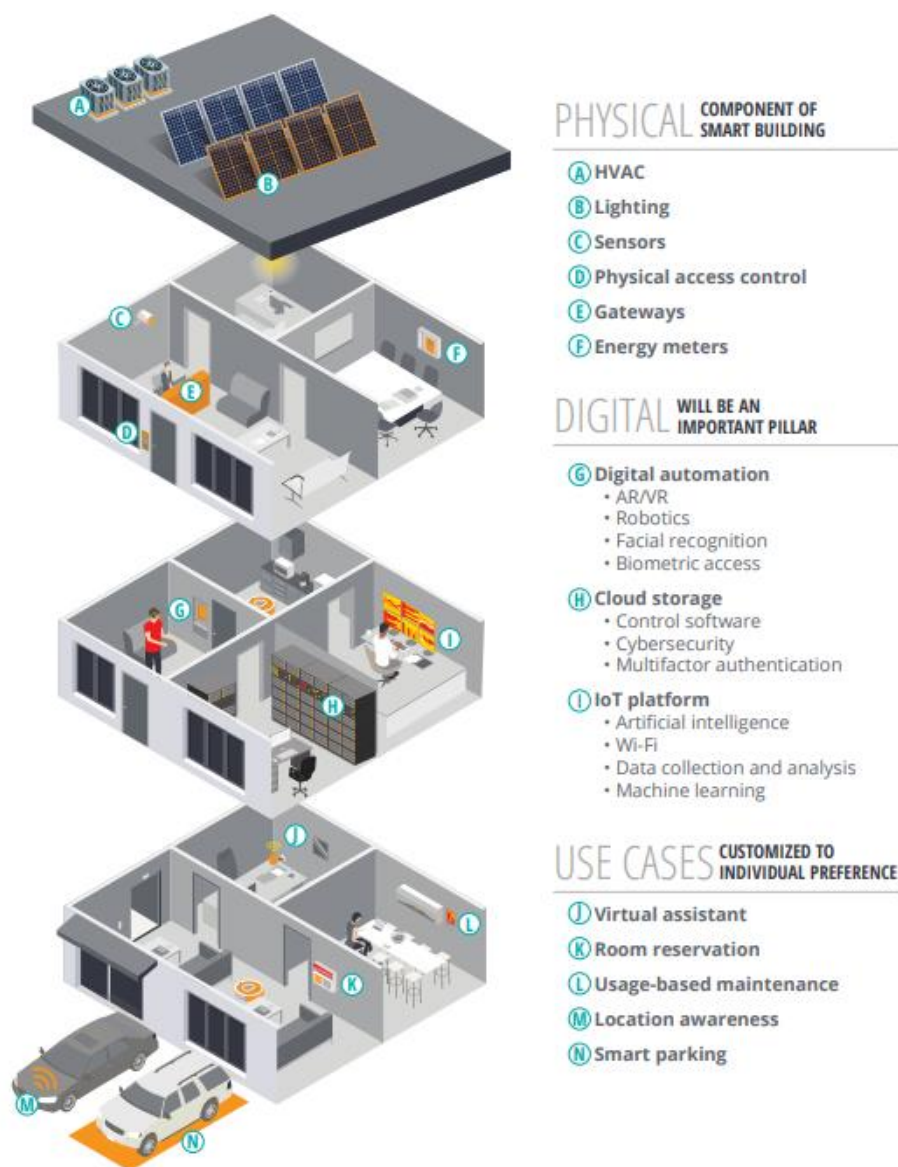


Figure 94: A comprehensive building optimisation ecosystem. Source: Deloitte, smart buildings – four considerations for creating people-centred smart, digital workspaces, 2018

In this view, smart buildings may represent a game-changer as, incorporating and integrating different technologies (e.g., IoT, Artificial Intelligence, edge and cloud computing), it can really make a difference in proposing an energy efficient, health safe & sustainable building model. Indeed, the provision of devices and technologies implementing the smart building paradigm can lead to better energy efficiency, improved occupant experience, and lower operational costs. Indeed, with the help of sensors measuring various parameters such as temperature, humidity, occupancy, energy usage, key card readers, parking space occupancy, fire, smoke, flood, security, elevators, and air quality, it is possible to get advanced data insight and to create a technological environment that efficiently and smoothly communicates with humans and vice-versa¹¹¹.

To implement such a system, the provision of a meta operating system for the IoT edge-cloud continuum will play a pivotal role, as it can avoid disadvantages of both the edge computing, that do not provide enough capacity for meta operating system for the IoT edge-cloud continuum, and the cloud computing, that may give unsatisfactory real-time responses. In addition, edge computing may also contribute to match data privacy requirements by not centralising data to foreign cloud data centres and providing an IoT environment closer to the end user.

¹¹¹ [Industry IoT Consortium, Intelligent Edge-Fog Architectures for Smart Buildings, 2022](#)

Europe is currently at a disadvantage on this front: in fact, North America dominates the market with 40% of installed base and is the most innovative in the sector, while Europe is forecasted to reach the 24% of the market in 2025¹¹².

4.5.4.1. Current trends and problems

The main driver of the smart building market growth will be related to the enhanced efficiency deriving from its implementation. First of all, smart buildings enhance energy efficiency, reducing electricity consumption thanks to the employment of advanced technologies that monitor the energy used by the different devices and appliances (e.g., smart lighting, smart heating, ventilation, and air conditioning (HVAC)). The occupancy indicator, which shows energy loss from unoccupied space, is another element that improves energy efficiency. In addition to this, smart building increases safety and security efficiency, allowing a better secure access control and alerting the energy centre in case of unauthorised intrusions. Furthermore, it can also detect anomalies such as fire. Finally, smart building solutions enhance employees productivity, making the workplace comfortable and more enjoyable, for example providing the right humidity level or monitoring and improving air quality¹¹³

Barriers to the proper development of smart buildings sector¹¹⁴ are listed below.

1. Absence of a common framework: it is needed to form a widely accepted understanding of the smart building framework and to set clear user requirements.
2. Expertise gap: the fast-growing environment of the smart building sector leaves few opportunities to have an in-depth overview of the field.
3. Lack of an end-to-end perspective: there is not a comprehensive view from the relevant stakeholders and operators.
4. High complexity of implementation and operations: smart building implementation and running is by its own nature very complex and is likely to intimidate potential stakeholders.

All these barriers are, in any case, offset by the value smart buildings can bring, both by differentiating commercial real estate operators from competitors and by generating sources of revenues. It is indeed estimates that smart buildings can provide¹¹⁵:

- 2 -17 % increase in resale value;
- 8 – 35% increased rental rates;
- 9 – 18% higher occupancy rates;
- 30% lower operating expenses;
- 9% higher net operating income.

4.5.5. Machinery of agriculture, forestry and construction

The agricultural machinery market is composed of several segments (e.g., tractors, ploughing and cultivating machinery, planting machinery, irrigation machinery, harvesting machinery). Because of the impact of the COVID-19 pandemic, this industry required more flexibility for machines to be fitted with transition engines already manufactured and procured compared to the statues before the crisis: in fact, the market slowed during the pandemic, due to global supply chains disruption as a consequence of government measures to prevent the spread of the virus. As of 2018, farm mechanisation in developing countries such as India and China accounts for 45-55%, whereas it accounted for 95% in developed countries such as the United States. Tractors accounted for over 43.6% of the market share in the agricultural machinery market in 2020. The Asia-Pacific region is expected to grow rapidly due to the high demand for agricultural products, mainly from India and China, over the forecast period. Furthermore, the market for farm machinery is anticipated to grow in Africa, since African

¹¹² [European Commission, Digital Transformation Monitor Smart Building: Energy efficiency Application, 2017](#)

¹¹³ Ibidem

¹¹⁴ [Deloitte, The future of smart buildings Six market insights on how to match expectations between occupiers and owners, 2022](#)

¹¹⁵ [European Commission, Digital Transformation Monitor Smart Building: Energy efficiency Application, 2017](#)

farmers are strongly requiring access to the latest farm technologies to enhance their farm operations¹¹⁶. The market in the next years will need to reach higher levels of productivity in order to meet an increasing demand for food since the United Nations predicts that the world population will grow from 7.7 billion people today to 9.7 billion in 2030 and even 11 billion in 2100¹¹⁷. It is, therefore, projected that agriculture IoT market will reach \$22.6 billion value by 2028 growing at a CAGR of 10.8% in the 2021-2028 time period¹¹⁸.

Global Forestry Machinery Market is usually divided on the basis of machinery types (e.g., skidders, forwarders, swing machines, bunchers, harvesters, loaders). This market is projected to continue its growth at a CAGR of 4.2% during the five-year period of 2020 – 2025. With the growing awareness of forest preservation and management. Combined with the increasing usage of machinery, has created an increasing demand for forestry equipment: moreover, rapid growth in demand for wood and wood-based products has led to the need for mechanised tree felling, creating opportunities for forestry equipment. Europe is the region who has been growing at the highest CAGR from 2021, and in perspective of the coming years up to 2026. Currently, it represents the largest market and is expected to remain the leading as well as the fastest growing market for forestry equipment. Financial support by the common agricultural policy (CAP) to rural areas, and the measures taken by the EU countries in order to encourage the forestry activities with the help of national development programs, continue to maintain the sales of forestry machinery in the region. Growing food demand globally has resulted in aggressive cultivation activities, which in turn has led to the conversion of forest lands into arable lands. This has further resulted in adoption of mechanised practices, which has created a significant demand for forestry machinery globally¹¹⁹.

The construction machinery market, comprising earth moving machinery and material handling, is consistently growing over the last years. In fact, the increased focus on infrastructure and automation in the construction and manufacturing processes provoked a significant impact on the construction machinery global demand over the past years, starting from 2017. Moreover, the improving economy and increasing construction activities across developing countries, like India and South Africa, are driving the potential demand for construction machinery¹²⁰.

The digital transition in farming, forestry and construction could bring several advantages for what regards an improvement in productivity and quality of the output, but also for farming, in the yield increase. Digitalisation also allows to have an integrated control and concertation of the machines (e.g., tractors, implements, combines in farming, pavers, rollers and trucks in road building, or forest harvesters and forwarders in forestry), production systems, sensors and devices, and so on. In particular, there is an increasing deployment of precision operations, in farming known as precision farming.

In these settings, most of the times located in rural areas, edge computing, coupled with temporary networks connections, will serve as an enabler allowing to bring intelligence to the agriculture, forestry and construction sectors.

Smart machineries can automatise several activities making the work more efficient, for example, for what regards forestry with the robotisation of silvicultural machines, equipped with instrumentation and autonomous motion control for the boom in the sector, as well as machine vision and laser scanners, that can perform a plant detection, distinguishing, for instance, between young and old trees, and deciding accordingly which need to be cut and which do not, carrying out an automatic point cleaning. Other machines can, through machine vision and LiDARs, improve productivity and operability, providing semi-autonomous machines that could also be used to update the forest information systems. The image below gives an idea of how smart forestry can be spread through the whole value chain.

¹¹⁶ Mordor Intelligence, “Agricultural Machinery Market - Growth, Trends, Covid-19 Impact and Forecast (2022-2027)”, 2021 [Agricultural Machinery Market Size, Share, Trends \(2022-27\) | Industry Forecast \(mordorintelligence.com\)](https://www.mordorintelligence.com/industry-reports/agricultural-machinery-market)

¹¹⁷ [United Nations, Population, 2022](https://www.un.org/en/development/desa/population/)

¹¹⁸ [Centre for the Promotion of Imports from developing countries \(CBI\) - The Netherlands Ministry of Foreign Affairs, The European market potential for \(Industrial\) Internet of Things, 2022](https://www.cbi.nl/en/insights/industry-internet-of-things)

¹¹⁹ Mordor Intelligence, “Forestry Machinery Market – Growth, Trends, Covid19 Impact and Forecasts (2022-2027)”, 2021. [Forestry Machinery Market Size, Share | 2022 - 27 | Industry Report \(mordorintelligence.com\)](https://www.mordorintelligence.com/industry-reports/forestry-machinery-market)

¹²⁰ Mordor Intelligence, “Construction Machinery Tires Market – Growth, Trends, Covid19 Impact and Forecasts (2022-2027)”, 2021 [Construction Machinery Tires Market Size, Share, Forecast 2022 - 27 \(mordorintelligence.com\)](https://www.mordorintelligence.com/industry-reports/construction-machinery-tires-market)

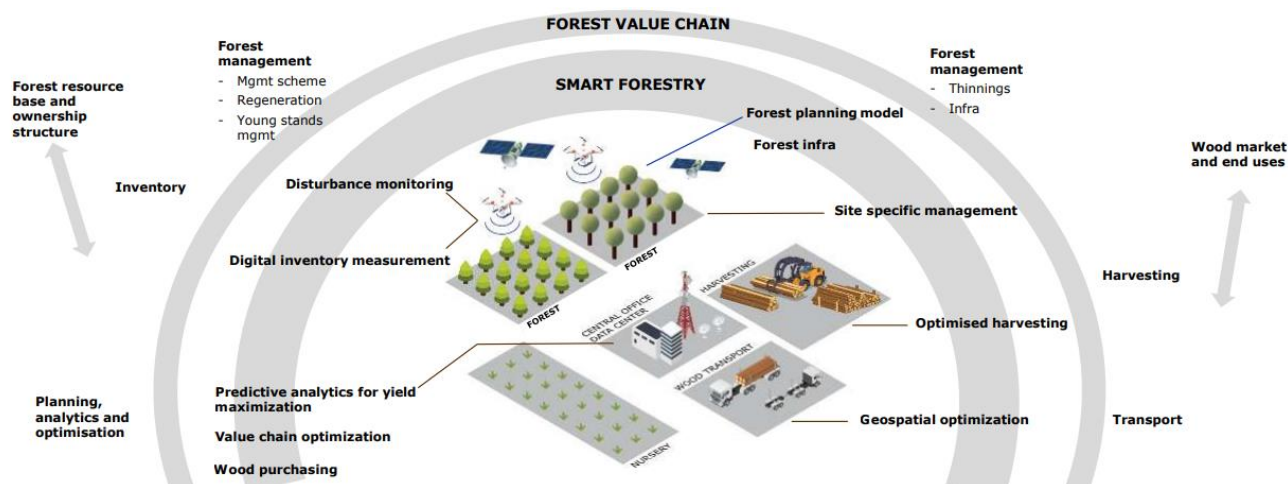


Figure 95: Smart forestry application. Source: Business Finland – Mediabank, Finnish solutions for smart forestry

4.5.5.1. Current trends and problems

Improved efficiency and the effort to reduce Greenhouse gasses emissions have been the priorities for the machinery industry over the last 50 years. For this reason, construction, farming and forestry machineries are heavily investing on the development of machinery equipped with cutting-edge smart technologies, capable of increase productivity and provide significant enhancement in terms of energy efficiency. Indeed, the improvement of cost and energy efficiency of these type of machines do not depend solely on the machine efficiency itself but is the result of the energy savings delivered through all the phases of the operations (e.g., drilling, cutting, collecting, pulling, ploughing, seeding, spraying, harvesting or transport). In this regard, the employment of ICT technologies, such as GPS, IoT, edge and cloud computing, in these machines is fundamental as they are able to process and integrate a large number of parameters (e.g., from weather and soil conditions to size and shape of the construction site) to reach higher levels of accuracy, for example in precision seeding and road placement.¹²¹ The adoption and investment on developing innovative and smart machineries, for example to make workers more efficient in remote controlling operations, represent the most promising tools to optimise agriculture, construction and forestry processes. Indeed, the aforementioned provisions and the shift to a process optimisation approach are making it possible to reach a series of gains, as intelligent machines can decrease the number of hours spent in merging processes together and adapting the tasks to the needs, reducing their number. In addition, connected vehicles allow to share data among machineries and to maximise the usage of machine park, by coordinating the activities on the farm, forestry and construction sites and by enhancing the planning of the needed activities to be performed¹²². These sectors are fully aware of the importance of digitalisation, IoT applications and automation in their fields. In this regard, McKinsey surveyed 400 senior executives of the construction sector. The survey confirmed this statement, as more than two-thirds of the interviewed executives said they believed industrialisation and digitisation will have the greatest impact of all upcoming upheavals in the business. Almost half of respondents believe that disruption will happen soon, within the next one to five years.

¹²¹ [Committee for European Construction Equipment \(CECE\) and European Agricultural Machinery Industry Association \(CEMA\), optimising our industry to reduce emissions, 2018](#)

¹²² Ibidem

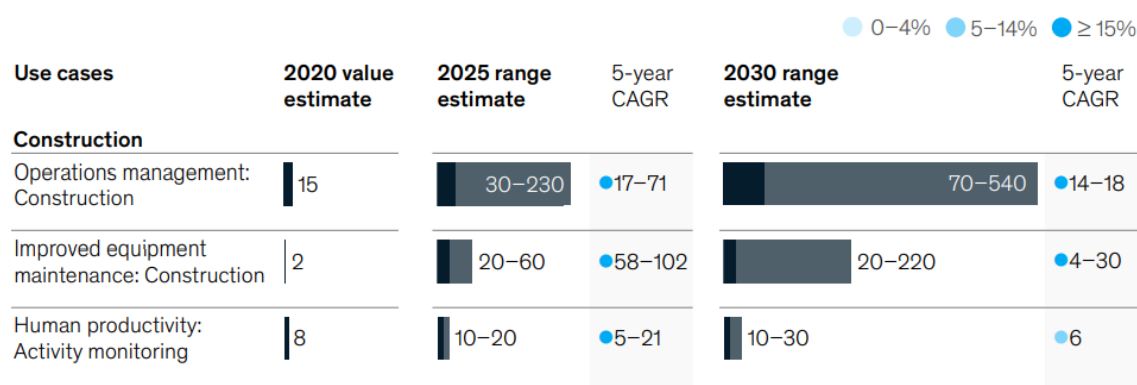


Figure 96: Estimated economic value by use case for construction, 2020–30, \$ billions. Source: McKinsey and Company, *The Internet of Things: Catching up to an accelerating opportunity*, 2021

An important application of IoT in construction industry will be operations management. In fact, the adoption of these technologies will increase, passing from 1-5% to 15-35% in 2030. McKinsey states that this use case could bring several advantages to the sector such as:

- increase in productivity by 5-10%;
- reduction of raw material cost by 5-9%;
- improvement in personnel efficiency by 7-15%.

This specific use case is projected to create an economic value of \$70 to \$540 billion annually by 2030. The largest part of this value will come from developed markets and China, that will account for the 44% and 39% of the total value respectively. Even though the emerging markets show lower rates of IoT adoption compared to the developed and to the Chinese market, these are expected to experience a marked increase in enabling technologies in the next few years, and to reach adoption rates comparable to the developed and to the Chinese markets.¹²³

Another important use case for the construction sector will be the improved equipment maintenance: indeed, machineries, such as bulldozers, cement mixers and cranes, are the fundamental part of construction industry. Nonetheless, these are inactive for a high proportion of time, amounting to the 36%: the ability to improve this data will be fundamental for construction companies to remain competitive on the market. IoT could provide a strong boost in improving this figure, with an increase in IoT adoption for this use case from the current 5% to the foreseen 25-40% in 2030, it promises to improve uptime by 30-50%, also enhancing productivity by 1-5%. This fact will translate to an economic impact of \$60-\$210 billion by 2030. In this case, the major contributors will be once again developed markets, accounting for the 30% of the total value, together with China, accounting for the 20%. The degree of implementation of these solutions will have a significant impact on the financial performance improvement brought about by the use of IoT applications in the sector. For example, companies with 50% of their fleets currently connected outperform competitors with lower adoption rates by 23%, while operators with 75% of their fleets connected have an improved performance of 51%.¹²⁴

¹²³ McKinsey and Company, *The Internet of Things: Catching up to an accelerating opportunity*, 2021

¹²⁴ Ibidem

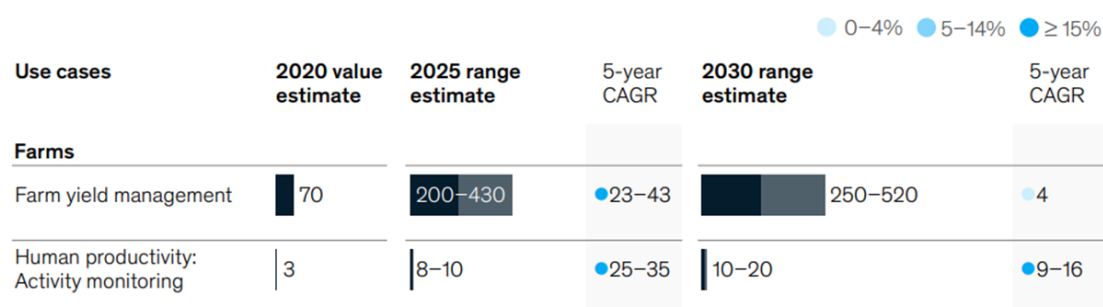


Figure 97: Estimated economic value by use case for farms, 2020–30, \$ billions. Source: McKinsey and Company, *The Internet of Things: Catching up to an accelerating opportunity*, 2021

The rise of IoT technology in farms strongly relies on a series of drivers, such as internet access: indeed, the number of IoT devices increased in the sector also in function of the higher internet access of farms that, for instance, in the US increased from 58% to 75% in the 2009-2019 time period. Other important factors are the lowered cost of technology, that decreased by more than 66% in since 2004, and the emergence of enabling technologies such as edge and cloud computing and Artificial Intelligence, that made the automation of the sector possible.¹²⁵

The main use case of IoT in the farming sector relates to the emergence of precision farming, that promises to provide an improved management of agricultural activities, using sensors and data to boost yield with potential increase of 15-20%. It is expected that the adoption of this kind of technologies could rise from 20% to 60-85% in developed countries. This could bring an economic impact of \$250-\$520 billion in 2030.¹²⁶

All of these automation and digitalisation features will have a lower impact on CO₂ emissions, thanks to the employment of edge computing technologies, that reduce latency and reaction time and, combined with 5G networks or Time Sensitive Networking (TSN) providing additional low latency, are able to reduce energy consumption when transferring AI and real-time embedded analytics.

The technological solutions described above could encounter difficulties to their implementation related to a series of perceived barriers. The most relevant ones relate to the potential risk of cyber-attacks, high up-front investment costs to acquire and implement the technological solutions, skill and professional shortages and lack of proper internet connection in rural areas, as well as an amount of time needed to process data that is perceived as too high. These barriers will be soon overcome by both a growing technological development, that will make it smoother to adopt such technologies, and by the generational change in the operators, that will be more willing and accustomed to emerging technologies.

4.6. Partners and Stakeholders engagement activities

Before conducting the interviews and focus groups, the interviewees were provided with a sample of the questions list that were going to be addressed during the interviews. They were also informed that their answers will be served to inform the project user requirements definition, design revisions and technology development, and that the anonymised summarised information was going to be submitted to the EC as part of public reports and, potentially, be used to write articles for peer-reviewed journals and relevant industry magazines, for presentations at conferences and workshops, and in the promotion of the project in general.

Furthermore, the interviews and focus groups have been recorded for the sole purpose of ensuring correct and timely information gathering. The interviewees agreed to be recorded at the beginning of each interview.

A series of peer-to-peer interviews were conducted with stakeholders and experts from the aerOS Consortium. The analysis activity focused on three different types of investigation. For each investigation, aerOS partners were selected on the basis of their ability to provide information in line with the scope of the activity.

Below are three paragraphs that will summarise the outcome of the three activities described above.

¹²⁵ Ibidem







¹²⁶ Ibidem

What emerged in the different interviews is reported in this order:

- Interviews:
 - general aspects;
 - economic aspects;
 - technological aspects.
- Focus Groups:
 - General aspects;
 - Economic aspects;
 - Technological aspects.
- Written Interviews:
 - Legal aspects;
 - Political aspects;
 - Environmental aspects.

To facilitate the usability of the information gathered, the results of the interviews are listed in order of relevance and in bulleted list form.

The legend below identifies the different topics assessed during the stakeholder engagement activities.

General	
Economic	
Technical	
Legal	
Political	
Environmental	

4.6.1. Interviews

Seven single interviews (lasting about 30-40 minutes) were carried out with coordinators (Project Coordinator and Technical Coordinator) and Technical Leaders (TLs) concerning market and technological trends. The results of the seven interviews can be found in Appendix A.

4.6.2. Focus Groups

Five focus groups were conducted (lasting about 1 hour) with Research Partners and Industrial Partners (including Tech & Use cases) involved in the project pilots (Smart Building, Renewable Energy Sources, Manufacturing and Production, Port Continuum, Machinery for Agriculture, Forestry and Construction). The results of the five focus groups can be found in Appendix B.

4.6.3. Written Interviews

Seven written interviews were administered to several experts regarding legal, political and environmental aspects (the partners involved were selected on the basis of the specific skills of individual organisations, in general and towards the project). For the type and complexity of questions, we preferred to receive written and thoughtful answers rather than answers in an oral interview. The results of the five focus groups can be found in Appendix C.

4.6.4. Online survey

In addition to the interviews, the written interviews and the focus groups, an online survey has been conducted, with the aim of gathering relevant and diverse feedback from major stakeholders, regarding the core topics which will object of deep and careful research during the whole project lifespan. Furthermore, the intention has been to assess specifically the current adoption level and actual needs required for any further adoption of Edge-to-Cloud technologies. The survey, created with LimeSurvey and customised to the specific features of the

aerOS project, has been widespread across the network of stakeholders linked to the consortium: nevertheless, the participation remained completely voluntary and free. It has been published online from October the 31st 2022 until November the 24th 2022. The aerOS consortium has introduced the proper question list with a brief presentation of the project goals and features, and with a disclaimer regarding data privacy, security and anonymization of the answer to the questionnaire, committing to maintain the strictest confidentiality of the research records.

The questionnaire has been composed of both mandatory and optional questions, divided into six groups: besides the general category, the remaining five have been dedicated to the PESTLE factors, that is to say Political, Economic, Social, Technological, Legal and Environmental, yet following a logical order of correlation among questions. All the questions that compose the aerOS online survey are listed in the following paragraph.

The survey has been addressed both to aerOS consortium members and to external specialists and experts. Through aerOS social medias the survey's link has been spread out.

4.6.4.1. Questionnaire

GENERAL

1. What type of Entity are you part of?
- 1.1. If you selected Company, what is your Enterprise Dimension?
- 1.2. If you selected Company, what is your position inside your Enterprise?
2. What is your core business?
3. What is your business scope?
4. What is the level of digitalisation of your Company?
5. Are you aware of any Edge-to-Cloud solution currently available on the Market?
6. Is your organisation rooting part of Service Brochure on Cloud or Edge Provisioning?
- 6.1. If you selected Yes: which one?
- 6.2. If you selected No: why?
7. Which departments or organisation processes would benefit the most from a well deployed and fully functional Edge Computing solution?

ECONOMIC

8. In your country and business reality, how pervasive and widespread is the adoption of Artificial Intelligence (AI) technologies?
9. In your country and business reality, how pervasive and widespread is the adoption of blockchain technologies?
10. How do you think that blockchain technologies can help in the certification of data and data providers?
11. Which are the first names that come to your mind concerning potential providers of Cloud Services?
12. Which are the first names that come to your mind concerning potential providers of Internet of Things (IoT)?
13. Which are the first names that come to your mind concerning potential providers of Edge Computing?

LEGAL

14. What are the main concerns and/or challenges experienced by your organisation for the deployment of IoT and/or Edge Solutions?

POLITICAL

15. Do you think that future regulations addressing the energetic crisis will influence the choice of adopting Cloud and Edge Systems?

16. Do you think that current and/or future political relations at the international level can in any possible way influence the decision of European Companies to adopt European Clouds rather than non-European?

SOCIAL

17. Is the availability of a skilled workforce a major concern for the adoption of Edge Computing technologies?

TECHNOLOGICAL

18. Do you use data to make business and/or operational decisions?

19. As a data scientist, what is the action in which you spend the most time?

19.1. How is the data you mainly use in your analysis?

20. What are the three most important features you demand on a Data Storage System?

21. Do you have to liaise with several heterogeneous sources of information?

22. How often do you have to liaise with several heterogeneous sources of information?

23. Do you have to manage different profiles and accounts in order to access different ICT systems for fulfilling your daily duties?

23.1. If you selected Yes, which different profiles and accounts do you have to manage in order to access different ICT systems for fulfilling your daily duties?

24. Do you think that your Company is experiencing interoperability issues due to the availability of different heterogeneous sources of information (e.g., each stakeholder owns a different stack of technologies)?

24.1. If you selected Yes, does this have a relevant cost for your Company?

25. Do you think that there is a real necessity of real-time (zero latency) applications or services for fulfilling your real needs and requirements?

26. In case you have to exchange data with several stakeholders, is there any Single Source of Truth (SSOT) common and shared which gives you the opportunity to share information and improve KPIs?

26.1. Are all the departments in your Company aware of that Single Source of Truth (SSOT)?

27. Does your organisation have more than one geographical location where Computing might take place?

27.1. If you selected Yes, do you think that prediction models having success in one spot might help all or some of the others?

27.2. If you selected Yes, do you think that prediction models having success in one spot might help all or some of the others?

28. Is the current architecture of IoT devices (if available) deployed in your infrastructure suffering bottlenecks due to the high volume of data?

29. Do you think that the bottlenecks related to the current architecture of IoT device due to the high volume of data have a significant cost?

30. Would you be secure enough if the data of your company would be sent to a Cloud-Based Node (outside your network)?

31. From 0 to 10, how would you value a system that could ensure that no raw data travels outside your network but still allowing you to share intelligence from Cloud Locations to Local Premises? (Of course, in comparison to a system that could not comply with such requirements.)

32. Are the traditional interfaces of the applications and services deployed in your company comfortable enough for you and the rest of your staff (please, bear in mind all the possible roles already available in your Company)?

33. What types of devices do implement your Edge Infrastructure?

4.6.4.2. Report and Statistics

The survey has been filled in by a total amount of 129 participants, of which 51 answered to the whole questionnaire while 78 responded only to their applicable targeted questions. A general profile of the aerOS survey participants can be seen through the graphs below (in which N/A and No Answers options have not been considered): the majority of them works in Companies (mostly large and medium sized) and Research Entities (including Universities), covering predominantly positions related to the general field of the European Research & Innovation projects (both technical and managerial figures).

Most of the Companies happened to be in the Technologies & Software, Telecommunications and Manufacturing Businesses, and revealed to be mostly technological competitive, in line with their national and an international business scope (the latter is slightly predominant).



Figure 98: aerOS survey participants general characteristics. Source: Own elaboration

In conclusion of the general part of the survey, it has been reconfirmed the expectation according to which the three business areas that could benefit the most from a complete and functional solution such as aerOS are Technology & Innovation, Research & Development and Engineering Business Units.

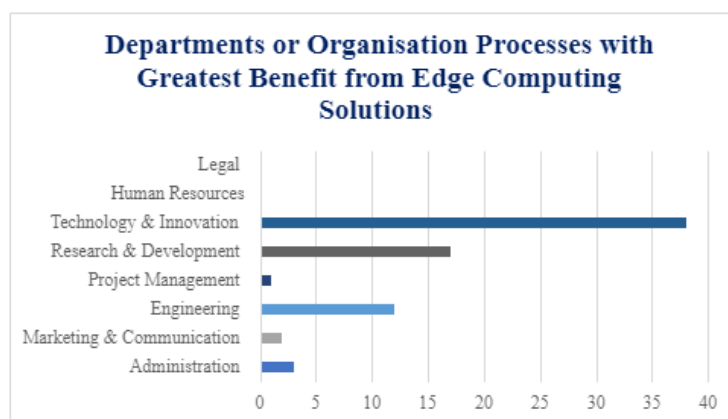


Figure 99: Business areas that could benefit the most from aerOS and related solutions. Source: Own elaboration

After learning that more than a half of the participants are currently aware of the availability of Edge-to-Cloud solutions on the present time market, at the same time the survey showed that only the 30% of them is actually rooting part of a Service Brochure on Cloud or an Edge Provisioning, and practically just the well renowned Microsoft Azure, Amazon Web Service IoT Core and some specific governmental CPDs have been mentioned, while the majority showed an almost complete unawareness of the existence of the aforementioned services.



Figure 100: Awareness of EtC solutions on the market; Position towards Service Brochure on Cloud and Edge Provisioning. Source: Own elaboration

Moving to the Economic side of the questionnaire, the level of diffusion of Artificial Intelligence technologies throughout the countries and/or business realities of the participants has been perceived as equally meeting and below the standards from the majority, immediately followed by above standards: the two opposite extremes, the far below and the far above levels compared to standards have revealed themselves as both almost irrelevant. Regarding the diffusion of blockchain technologies, for the largest part it resulted to be below standards, with a few cases of far below and meeting standards, while just a handful of far above the average. Despite these results, blockchain technologies are significantly perceived as pivotal in the certification of both data and data providers. Concerning the first names that that came to the participants' mind with regard to potential providers of Cloud Services, the following have been mentioned: Amazon Web Service, Microsoft Azure, GoogleCloud, Aruba, FastWeb and IBM Cloud Service. Microsoft Azure, GoogleCloud, IBM and Amazon Web Service returned also as main IoT providers, plus a few mentions for Siemens, Apple, Android Huawuei, Cisco, ABB, PTC, Softeq, Bosch, Raspberry Pi, and Arduino. The majority of the aforementioned providers won the "contest" as best renowned Edge Computing providers, together with NEC, ORACLE, DELL, Schneider, Intel, Edge Impulse, Deutsche Telekom, and EdgeConneX.

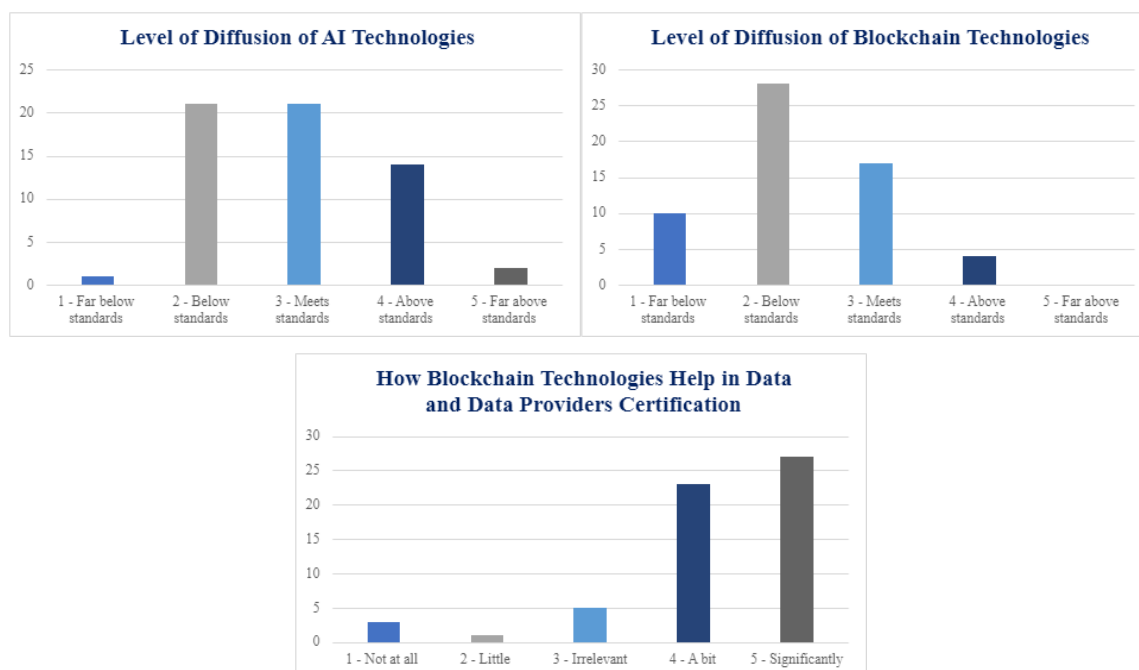


Figure 101: Level of diffusion of AI and blockchain technologies, and the latter estimated improvements in data and data providers certification. Source: Own elaboration

From a legal point of view, the main interest of the aerOS survey has been to discover which are the main concerns and challenges usually experienced by companies in the deployment of both IoT and Edge solutions: starting from the fact that all the supposed challenges have proven to be met in a certain measure, the majority of answers referred to the complexity for devices integration, followed by data collection and analysis, privacy, security issues, scalability, but some mentions went to vendor lock-in and maintenance costs.

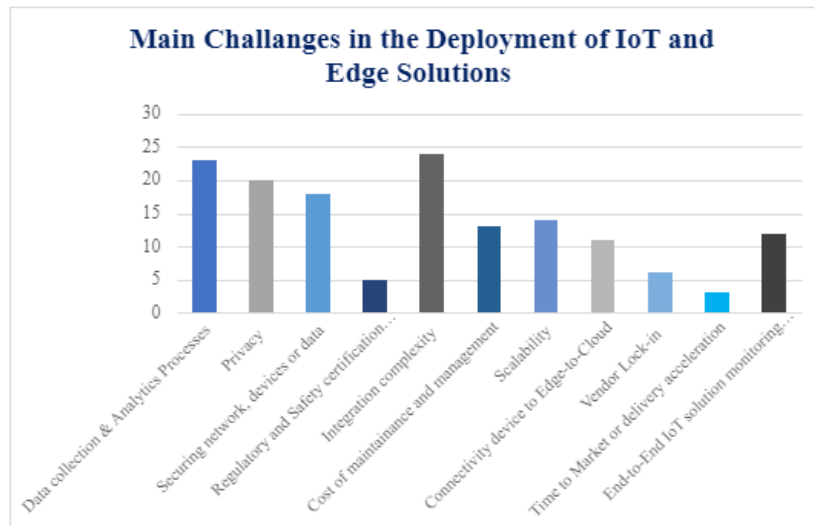


Figure 102: Major concerns and challenges regarding IoT and Edge solutions deployment. Source: Own elaboration

One of the main political concerns treated by the survey has been the potential impact of future regulations regarding the Energetic Crisis on the adoption of Cloud and Edge Systems by the EU companies: though the majority of participants did not perceive the aforementioned as a too critical discriminating element, several responded with their significant perception of the impact. Furthermore, most of the interviewed proved to give to international political relations on the adoption of a European rather than a non-European Cloud a quite strong influence.

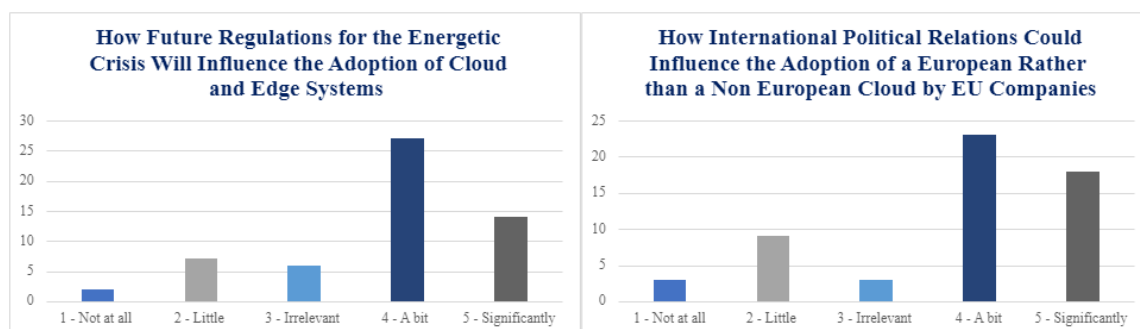


Figure 103: Perceived concerns of political issues related to the adoption of Cloud and Edge systems, and of a European vs Non-European Cloud by EU companies. Source: Own elaboration

The most significant social instance about which feedback from stakeholders was required has been the availability (or lack of it) of a properly trained and skilled workforce: the related influence that the phenomenon could have on the adoption of Edge Computing technologies has been perceived by participants as significant, or at least noticeable.

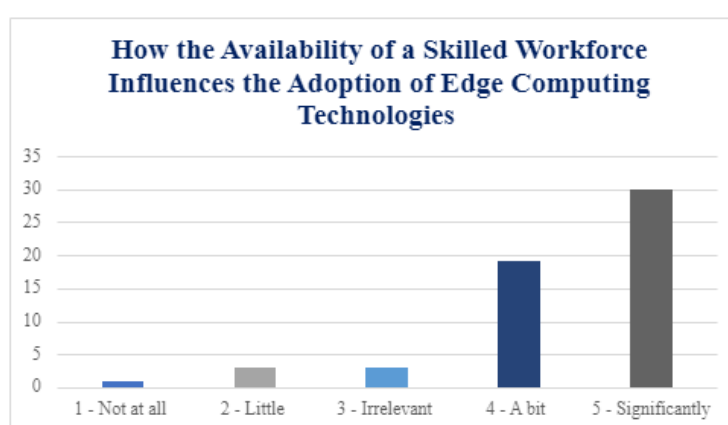
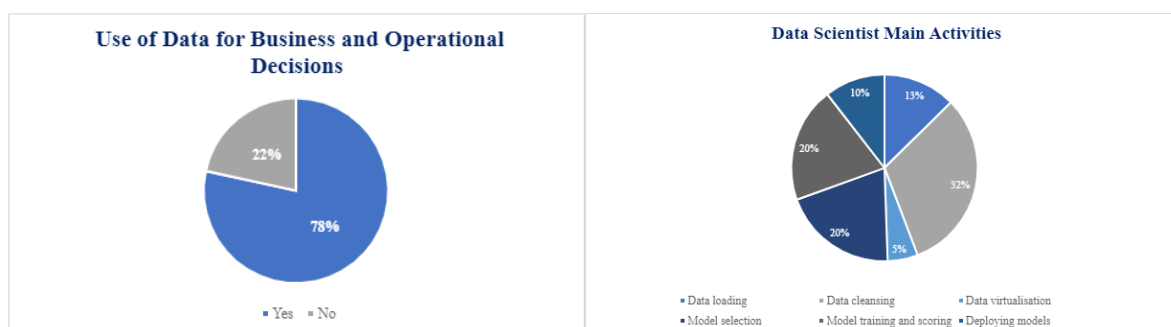


Figure 104: Influence of social factors like the availability of a skilled workforce on the adoption of Edge Computing technologies. Source: Own elaboration

Moving to the technological factors, the next group of questions concerned the use of data (mostly raw, excel, textual, csv) for the participants inside business operations: according to the answers gathered in the three graphs below, it has been observed that the 78% of participants use the data gathered from their IoT or ICT systems in order to make operational decisions in a daily or weekly basis. Since the undoubtable relevance of data, the aerOS consortium decided to focus on the specific activities on which data scientists spend the majority of their working time: the answers have covered all of the proposed options, especially data cleansing, data loading, and model selection. As for the perception of how should be the main characteristics of an ideal data storage system, all the features provided as possibility by the aerOS consortium have been judged as particularly relevant, with a special mention for data continuity and accessibility, effectiveness of the security, reliability of data preservation and quick recovery of lost material.



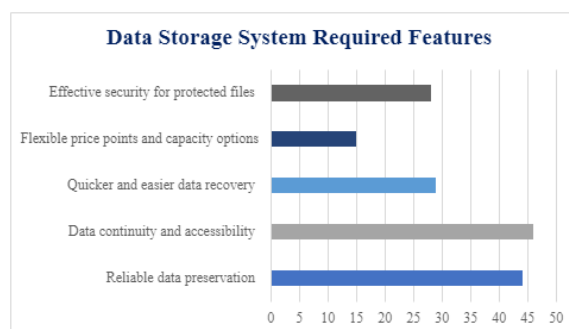


Figure 105: Data relevance in business and operational decisions; data scientist main activities; most required features for data storage systems. Source: Own elaboration

Considering that, as the graphs below are showing, almost the whole group of participants has to liaise with many different heterogeneous sources of information (often and sometimes have been the most frequent answers to the specific question on the temporal occurrence), it must be noticed how this particular phenomenon is directly connected with interoperability issues, experienced by almost the 70% of the survey fillers and with relevant costs as a consequence for the 54%. Moreover, more than a half of them has to manage different profiles and account related to different ICT systems in order to fulfil their daily duties.

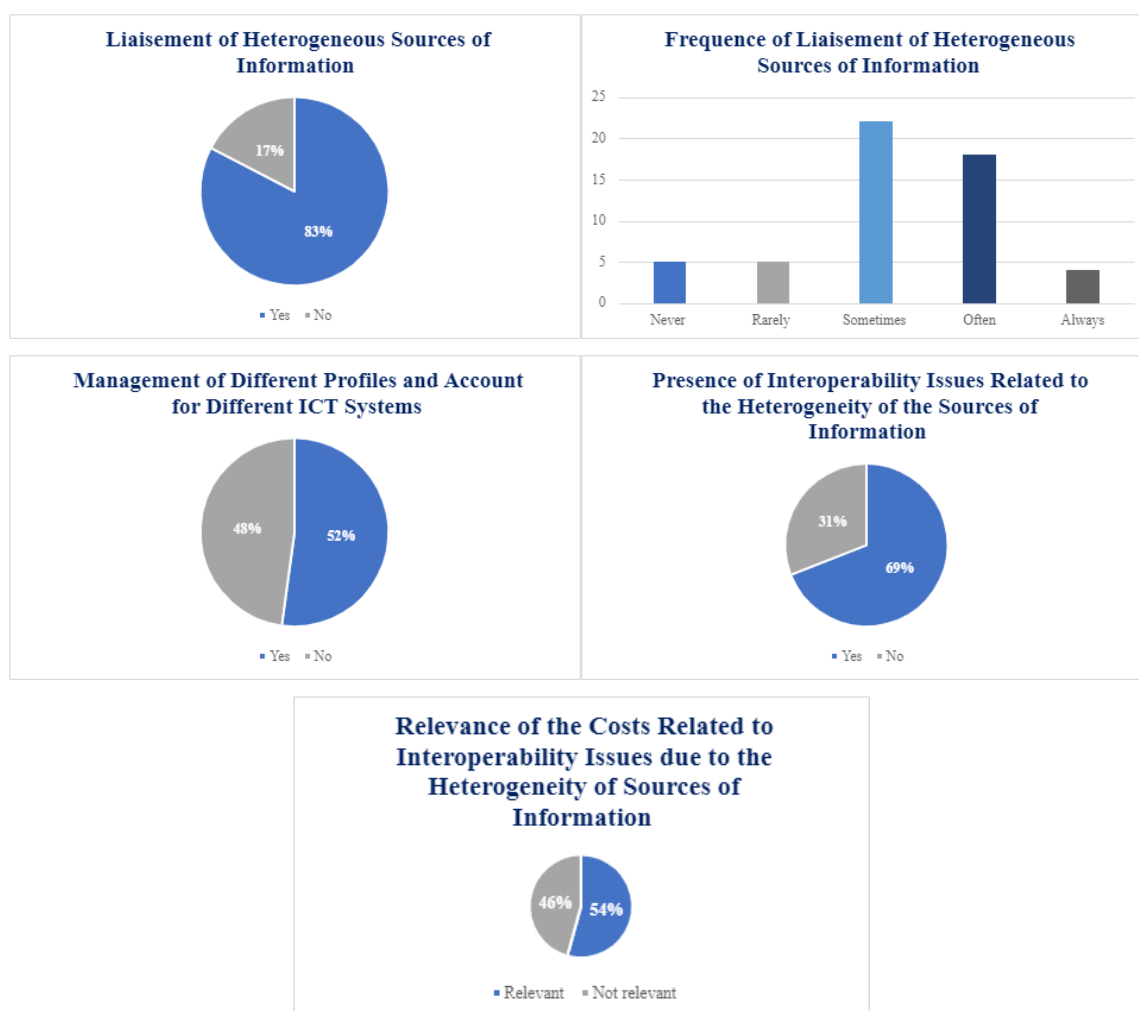


Figure 106: Technical aspects related to the presence of heterogeneous sources of information. Source: Own elaboration

Going forward, the generally acknowledged necessity for zero latency applications and services has been confirmed by what the participants answered on the matter, affirming the potentiality of a solution such as aerOS on the market. In case of necessity to exchange data with several stakeholders, the consortium was interested in

the use of any Single Source of Truth (SSOT) architecture in order to both share information and improve Key Performance Indicators on the performances: of the 32% of participants who replied positively, only the 23% happened to believe in the common acknowledgment of that SSOT by their own company.

Almost the 70% of the interviewed organisations revealed to have more than one geographical location where Computing is currently taking or might take place, showing a firm belief about the positive influence that a successful prediction model for computing used in one location could have on the others.

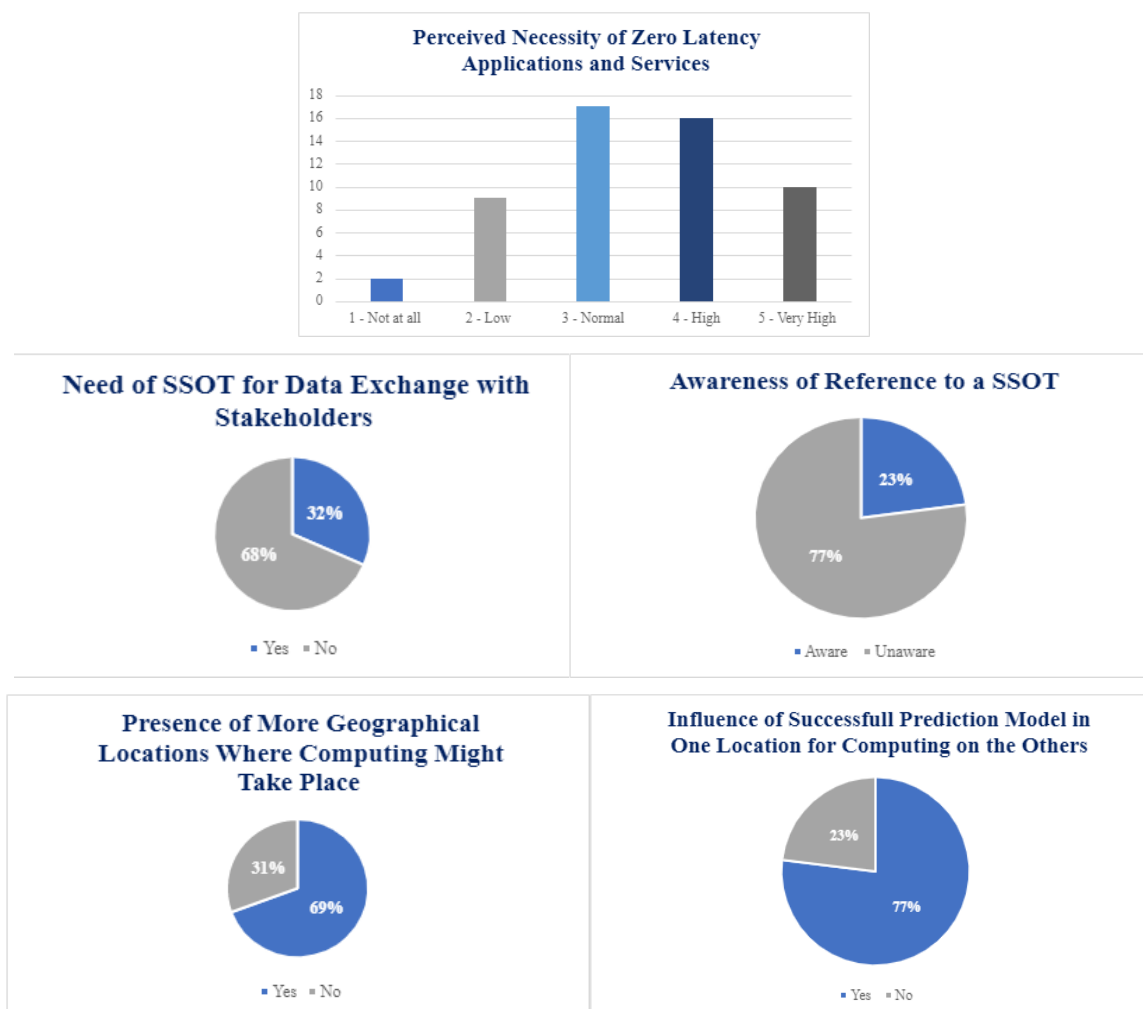


Figure 107: Perception of Zero Latency necessity; SSOT company reference; successful prediction model influence on correlative geographical location. Source: Own elaboration

As easily predictable, the aerOS survey participants recognised the importance of very strict privacy concerns for data sharing, at the same time pointing out that, for the 67%, high volumes of data do not necessary bring to bottlenecks for currents IoT devices architectures, and, even in the case of their occurrences, the cost is not perceived as too significant in general terms. Yet, regarding the security in letting company data travel throughout cloud-based nodes outside their respective networks, that is an argument that has divided the participants almost in half.

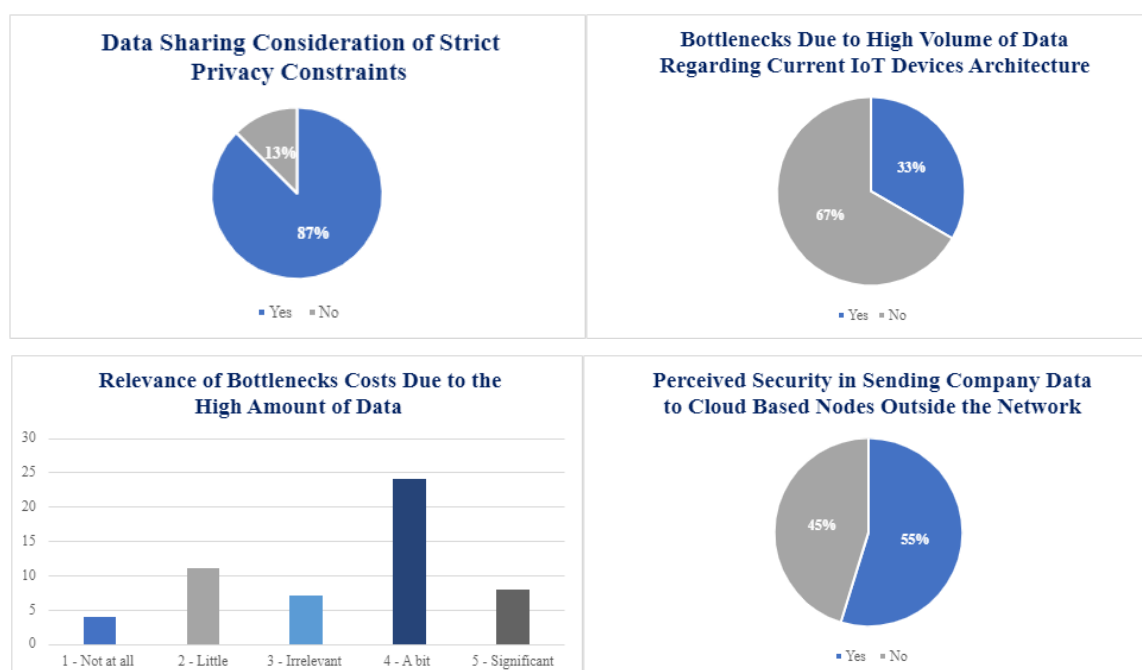


Figure 108: Privacy concerns for data sharing; bottlenecks related to huge amount of data; perceived security in letting data travel to cloud-based nodes outside the network. Source: Own elaboration

The highest rungs of a rating scale with 10 as maximum value have been awarded to systems that possess the ability to enable intelligence sharing from cloud to local premises while still avoiding the dispersion of data outside the network of their owners. Regarding the topic of comfort of the interfaces for applications traditionally deployed by the participants' companies towards their user-friendliness to the staff, the participants have been almost divided in half. In conclusion, as a last question, the aerOS consortium suggested a series of devices in order to understand which of them could better improve participants Edge infrastructure: all the options have been chosen by someone, with particular attention dedicated to sensors, servers and gateways.

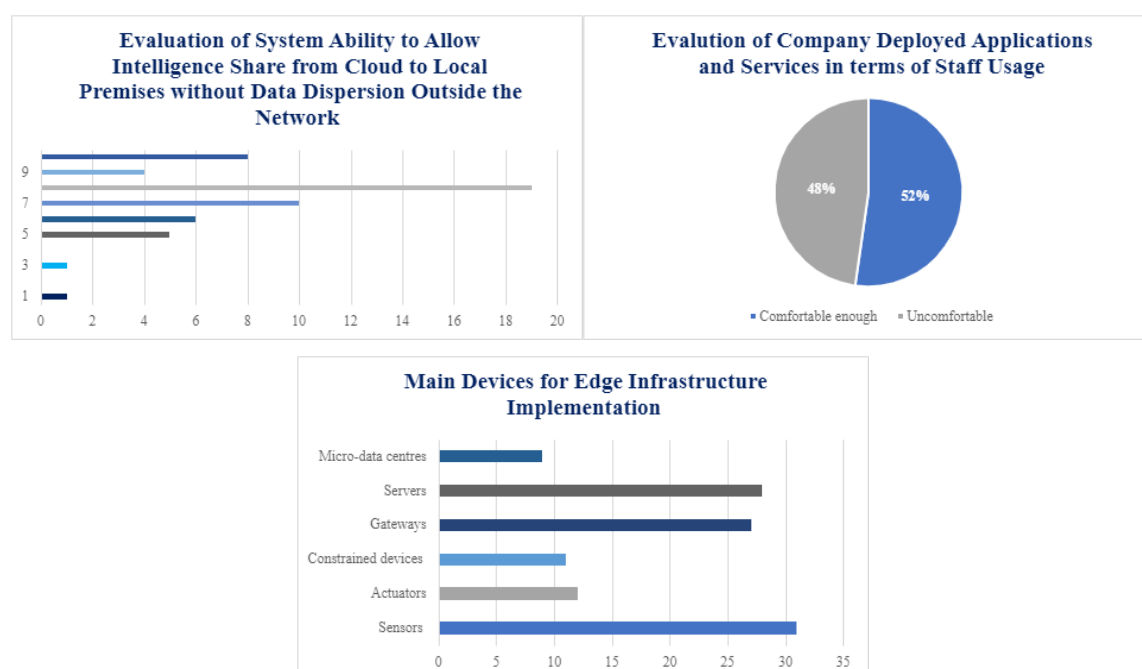


Figure 109: System Evaluation regarding intelligence from cloud to on premises; evaluation of company deployed applications and services based on their staff user friendliness; main devices that implement the edge infrastructure of the participant. Source: Own elaboration

4.6.5. Workshop

4.6.5.1. Reported Activities

The workshop, the final event of the Task 2.1 “State-of-the-Art and Market Analysis” of the aerOS project, has been held on November, the 29th 2022. It was a two hour online live event, moderated by DST and conducted using Microsoft Teams conferencing tool. 55 participants attended in total, coming from both industrial and academic stakeholder groups. In addition to the consortium members, key stakeholders were invited. So, the event was not aimed at the general public. Further, the workshop has been recorded for the purpose of ensuring correct and timely information gathering and to allow further offline feedback and reviews: all the participants agreed to be recorded and to the sharing of the material at the beginning of the event. The video of the workshop will be uploaded on YouTube.

As its pivotal goal, the final workshop aimed at getting further insights regarding the envisioned instances and necessity concerning Edge and Cloud Computing and the edge-to-cloud Continuum, in addition to the considerations that have been already drawn from the previous engagement process of the interviews and focus groups, as well as informing partners about the outcome of task 2.1: the scope has been successfully reached by welcoming different technological and industrial stakeholders that were able to provide a significant input and to exchange ideas with the project partners on the conceptualisation and formulation of the key messages of aerOS. Furthermore, relevant resources have been shared, and individual feedback for each participant has been overviewed, as already collected within this deliverable “D2.1 - State-of-the-Art and Market Analysis Report” of the aerOS project. Furthermore, those insights will be considered throughout the whole project lifespan, as some of the questions raised will find a proper answer only after the solution actual development and the interaction with partners and stakeholders has been planned to continue after the conclusion of the Task 2.1.

This workshop has been planned and designed by the aerOS consortium with some specific objectives in mind, that is to say:

- i. to introduce a general overview of aerOS framework, unique features and selling points, identified thanks to the research conducted by the consortium and to the continuous integration between partners and stakeholders in the aforementioned engagement process carried out through interviews and focus groups;
- ii. to answer the questions made by participants and therefore to set up an open discussion with industrial and academic partners finalised at both confirming and summarising the previously obtained feedback, and sharing, in a mutual and fruitful way, the most significant needs, requirements and real-live experiences or pains which could affect the aerOS staff, before the phase of definition in detail concerning verticals and use cases.

The two-hour event has been divided into two parts.

1. Report on the results of the activities planned in T2.1.
2. Q&A and open discussion.

During the first part, UPV and DST have been the speakers, while all questions, that arose both previously and during the listening, have been collected in the chat and answered, in order, in the Q&A session. Regarding the open discussion session, it has been moderated by DST, who let the participants speak by a show of hands.

The detailed agenda for the workshop is described below:

- **Workshop Introduction:** the Task Leader and Moderator DST welcomed the participants and performed a general introduction of audience to the workshop itself, its modality and its agenda.
- **Project Overview:** the Project Coordinator UPV, described briefly the aerOS project, focusing on objectives, testbeds, and timeframe.
- **Task 2.1 Description:** DST proceeded at describing the specific goal of the Task 2.1 - State-of-the-Art and Market Analysis and the content elaborated in the report that constitutes the Deliverable 2.1.
- **State of the Art Recollection:** UPV concentrated specifically in the description of the State of the Art concerning Technology.

- **Market Analysis Focus:** DST performed a general overview of the Market Analysis developed and carried out by them.
- **Interviews, Focus Groups and Online Survey Results Overview:** DST presented the results of all the previously conducted engagement activities, that is to say: interviews (both oral and written), focus groups, and the aerOS online survey.

Questions & Answers: questions from the audience have been addressed, and the open discussion has been carried out before the workshop conclusion.

4.6.5.2. Report

Mr. Andrea Valerio Chentrens from DST kicked-off the workshop presenting the workshop modality, the agenda, and asking for permission for recording. Then he left the floor to Mr. Ignacio Lacalle Úbeda from UPV, who provided an overview of aerOS, highlighting that the project aims to overcome the lack of capabilities to handle the new requirements foreseen in the field of Edge and Cloud Computing and their Continuum. The project has identified several use cases or vertical scenarios that will be demonstrated by means of the architecture of aerOS in five different pilots: Manufacturing and Production, Renewable Energy Sources, Smart Building, Port Continuum, and Machinery for Agriculture, Forestry and Construction. It has been pointed out how at least two pilots will be added during the second phase of the aerOS project thanks to two open calls, meaning Automation, Transport & Mobility, and Health. Furthermore, several key points of the project from an innovation and exploitation point of view have been mentioned, and among them:

1. the project not only wants to develop and deploy the state-of-the-art technologies, but also to go beyond, as the project is a Horizon Europe Research and Innovation Action (RIA), which also demands to carry out scientific research beyond market solutions;
2. all the solutions to be used will have as a key pillar a human centricity approach;
3. the project goal is to solve real problems and concerns from the specific industrial domains: that has been the reason behind the choice made by the consortium to involve most relevant stakeholders of those domains. It has been also reminded that the online workshop has been set up in order to get the thoughts and to extract opinions of different experts and stakeholders across the five pilot environments of the project.

Afterwards, Mr. Chentrens from DST explained the specific goal of the Task 2.1 - State-of-the-Art and Market Analysis, focusing on content elaborated in the report that constitutes the Deliverable 2.1, before leaving the floor to Mr. Lacalle Úbeda from UPV, who proceeded in presenting the current State-of-the-Art of Technology in each of the chosen pilots and testbeds, while also explaining why the latest advancements are not enough to fulfil all the industrial and business needs. More in detailed, the following topics have been quickly but effectively summarised, as they have been addressed in the D2.1. Report:

- Edge-to-cloud continuum orchestration.
- Smart networking and infrastructure management.
- Resource orchestration approach.
- APIs, monitoring and communication services for the continuum.
- Data orchestration approaches.
- Review of relevant techniques for the meta operating system.
- Containerisation and virtualisation techniques.
- Edge-native approaches: cloud-native techniques applied along the computing continuum.
- Self-* capabilities of heterogeneous nodes.
- Data syntactic and semantic interoperability in the continuum.
- Data sovereignty, governance and lineage policies.
- Advanced AI management approaches.
- Security, integrity, trust, privacy and policy enforcement in the computing continuum.
- From DevOps to DevSecOps to DevPrivSecOps.
- Distributed multiplane analytics.

- Surrounding ecosystem overview.
- Industrial approach to edge-to-cloud continuum in Industry (I4.0 and I5.0).
- Current existing standards related to aerOS.
- Review of the DATA-01-05 cluster.
- Other related projects.
- Review of current approaches in selected verticals.
- Edge-to-cloud technologies in robotics and manufacturing sector.
- Edge-to-cloud technologies in maritime port sector.
- Edge-to-cloud technologies in machinery construction sector.
- Edge-to-cloud technologies in telecom operators sector from a usability perspective).
- Edge-to-cloud technologies in renewable energy production.

Then, it has been the turn of describing the Market Analysis carefully conducted by Mrs. Sara Gaudino from DST, whose research and conceptualisation has touched all the main markets considered relevant for aerOS:

- The general aerOS market, divided into the target market, with a focus on the cloud computing and the edge computing specific segments, and the correlative market, composed of the IoT, the AI, the telecommunication, and the blockchain segments.
- The market size and growth for both edge and cloud computing.
- The most significant market trends and drivers for the aforementioned topics.
- The influencers market factors, divided according to the PESTLE methodology factor segmentation (Political, Economic, Social, Technological, Legal and Environmental).

At this point of the workshop, DST shifted the focus in particular on the partners and stakeholders engaging process, that was constituted by oral and written interviews, focus groups, dedicated to the feedback from the pilot verticals and the online survey. Mr. Chentrens showed how the whole path followed the factor segmentation proposed by the PESTLE methodology, plus a general overview: the response from partners and stakeholders regarding the technological and economic aspects of aerOS was analysed both through the oral interviews to coordinators and technical leaders and through the end users dedicated focus groups, while the legal, political and environmental approach of the project was dealt with through the written interviews to significant experts. Then, the results obtained through the aerOS online survey have been presented and analysed by Mrs. Gaudino. It has been remembered that also the questionnaire of the online survey has been created according to the PESTLE approach and administered to further external stakeholders in order to improve what was previously collected through the partners engagement in interviews and focus groups.

Afterwards, the second phase of the workshop started with a Question & Answer session, in which the audience raised their opinions or doubts about potential solutions to overcome the presented challenges, leading to an open discussion. The list included the following exchanges.

- **Regarding the data collected through the Market Analysis, in particular those about the differences between Europe and the United States, where it has been noticed how many more households in the US are covered by 5G compared to what happens in Europe: have those data been correlated to some geographical distribution for households?**

Considering that it is quite simple just to underline a difference among countries but that the mere fact is not significant on its own unless it is properly investigated what lies behind the gap, the first element that has been taken into account in the Analysis was the presence of divergent features for the respective markets. In fact, the US market is way more concentrated, whereas in Europe we have a fragmented market scenario, thus making the US player stronger and able to lower their costs and to invest in the development of technologies of sorts. The geographical position of the interested population, located in rural or urban areas inside the European continent may be a factor contributing: the results showed that the gap between urban and rural population coverage in Europe is extremely significant by itself. For example, with regards to 4G, the percentage of uncovered households was about the double in rural areas compared to the average number of uncovered households in general.

- **Regarding the two concepts of data fabric and data mesh, the aerOS research showed how they are actually compatible and able to live and work together: how will that be possible? What is the purpose of that, and how could it make sense?**

Data fabric and data mesh are in truth complementary, since the first is an enabler to the second: data fabric just designs and installs a way of how the data sources are exposed to data consumers, in terms of a graph that includes which data entities are related to which folder, which information is available in terms of attributes, metadata and all related aspects. The creation of a graph allows to consult queries and to be able to map on information in an easier way. Data fabrics provides all those elements, helped by tools that allow that interconnection of brokers, mostly. On the other hand, data mesh regards how that information is exploited once properly understood what is what or what does what or has in terms of attributes, and so on. So, the data mesh should allow to define the exploitation of those actors, focusing more on the data consumers, who they are, how they are federated, which are their rights to exploit the data and treat the data as a product: mesh can be considered as a philosophy of exploitation for the data, while fabric allows data to be understood in the different metadata and capabilities. It has to be pointed out, though, that the tools currently existing on the topic are not yet stable enough, thus making the line between data fabric and data mesh very thin, at least for now, and aerOS is investigating on that element, too.

- **How does the aerOS consortium plan to deal with the governance of all those data coming from the aforementioned two concepts of data fabric and data mesh also, in perspective of a general consideration of all players involved, not only Europe and United States but also, for example, Asia.**

The aerOS consortium includes some partners (e.g., Telefónica and Ericsson) who are currently working on the topic of data governance, how it is designed and envisioned, in close contact with the European Commission. Generally speaking, even though the matter is very pivotal for the project and will be properly dealt with in time, it can be said from now that the aerOS consortium will be basing on metadata inclusion every time a piece of data is inserted, thus making sure that, in the very moment an information gets into the system, it has a tax labelling indicating the origin of the data, their ownership, their time of creation and source, their access rights or access constraints.

- **The integration of legacy systems has been noticed to actually cut across some of the comments and issues about, for example, interoperability and heterogeneity of data, since those elements have been presented as current barriers to the market entrance: could the integration of legacy systems effectively play a significant role and strengthen the value of the aerOS solution?**

The aerOS consortium is not currently certain about the extension of the reach of legacy systems and demonstrated equipment in the project, since it depends mostly on the kind of pilots selected and brought in. Nevertheless, it is very important to get the knowledge of possible future synergies with those other projects which are currently dealing with legacy systems integration.

- **Is aerOS envisioned to bring some benefits to data scientists in particular?**

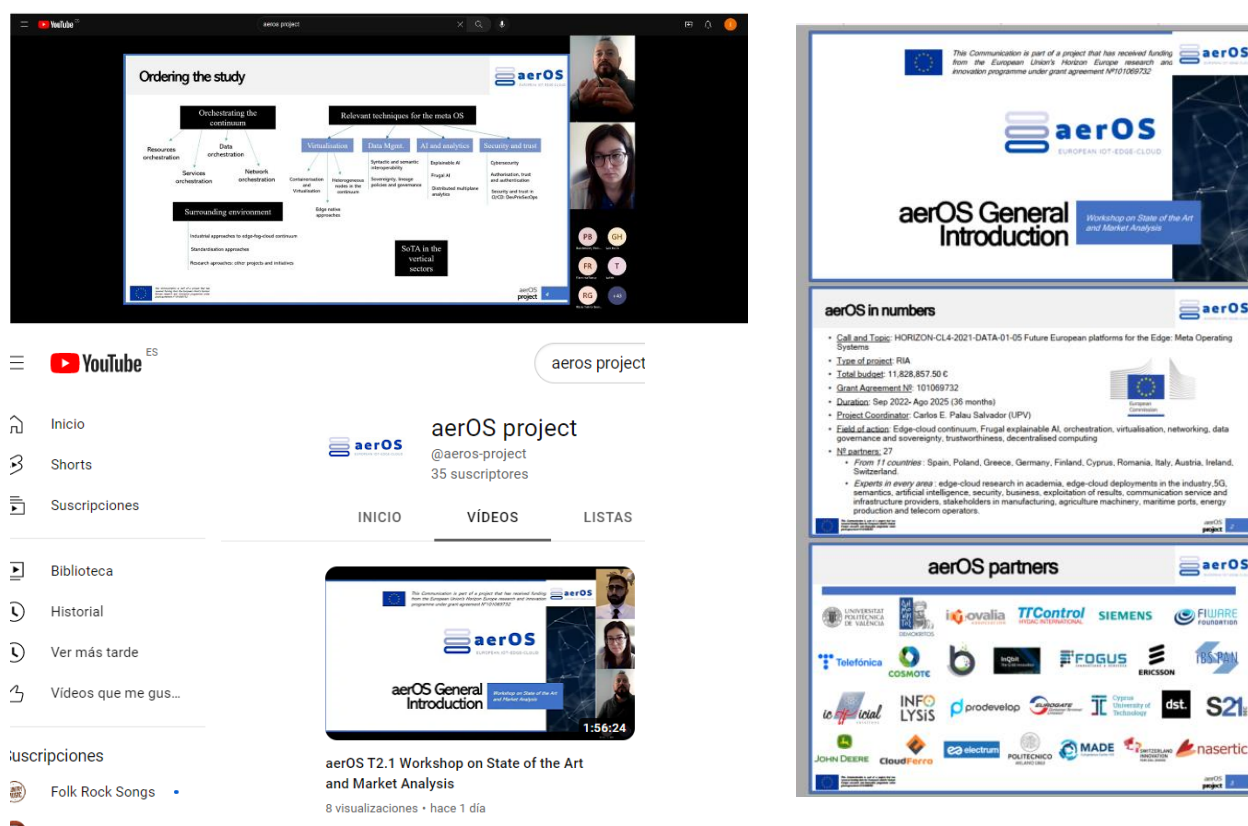
Compared to what the survey showed about the statistics of which activities currently require most of the data scientist time (e.g., labelling and pre-processing the data), aerOS aims at using a frugal AI approach, thus reducing the need for such a vast amount of data and decreasing the time spent by data scientist in dealing with and managing them, allowing to concentrate on more productive activities in their job.

5. Conclusions

This document has evidenced the results obtained from the activities committed in tas T2.1, which has been the main focus of the Consortium during its first three months of execution. After conducting intensive activities of systematic research and review of technologies, together with thorough analysis of the market related to aerOS outcomes, several conclusions are extracted:

- The variety of technological domains tackled by aerOS increases the challenge that it will face.
- Having properly structured the scope to research and advance upon is putting aerOS in the best position to overcome those challenges and exert the planned innovation.
- Orchestration can be mainly conceived from the network, service, resources and data in the continuum.
- The five pilots of the project are very well aware of the current available technologies and the paths that must be followed to meet their objectives
- aerOS is in a privileged position to build on top of the scientific findings exposed in the document.
- The market analysis clearly shows that the edge-cloud computing continuum is a niche where aerOS outcomes will fit perfectly.
- There are other research projects with similar or complementary goals to aerOS that will need to be observed and interacted with to ensure proper innovation, especially those within the DATA-01-05 cluster.

Relevant conclusions were exposed during the workshop held on November, 29th, 2022. The reader is kindly invited to gather more detailed information by watching the recording of the video and by consulting the presented slides in such event:



Ordering the study

- Orchestrating the continuum
 - Resources orchestration
 - Services orchestration
 - Network orchestration
 - Data orchestration
- Relevant techniques for the meta OS
 - Virtualisation
 - Data Mgmt.
 - AI and analytics
 - Security and trust
- Sampling environment
 - Industrial approaches to edge-cloud continuum
 - Standardisation approaches
 - Research approaches: other projects and initiatives
- SoTA in the vertical sectors

aerOS in numbers

- **Call and Topic:** HORIZON-CL4-2021-DATA-01-05 Future European platforms for the Edge: Meta Operating Systems
- **Type of project:** RIA
- **Total budget:** 11,828,857.50 €
- **Grant Agreement No:** 101069732
- **Duration:** Sep 2020- Ago 2025 (36 months)
- **Project Coordinator:** Carlos E. Palau-Salvador (UPV)
- **Field of action:** Edge-cloud continuum. Frugal explainable AI, orchestration, virtualisation, networking, data governance and sovereignty, trustworthiness, decentralised computing
- **NE partners:** 27
 - From 11 countries: Spain, Poland, Greece, Germany, Finland, Cyprus, Romania, Italy, Austria, Ireland, Switzerland
 - **Experts in every area:** edge-cloud research in academia, edge-cloud deployments in the industry, 5G, semantics, artificial intelligence, security, business, exploitation of results, communication service and infrastructure providers, stakeholders in manufacturing, agriculture machinery, maritime ports, energy production and telecom operators.

aerOS partners

UNIVERSITAT DE VALÈNCIA, ovalia, TTControl, SIEMENS, FIJURE, Telefónica, COSMOT, b, FOGUS, ERICSSON, ioficial, INFO LYSIS, prodevelop, S21, JOHN DEERE, CloudFerro, electron, POLITÉCNICO, MADE, nasartic

Figure 110: Evidences of the conclusions extracted out of the D2.1

References

Smart Network and Infrastructure Management

- [SNIM-1] ETSI, “Network Functions Virtualisation (NFV) - Network Operator Perspectives on Industry Progress,” White Paper, 2015.
- [SNIM-2] A. Tzanakaki, M. Anastasopoulos and I. Berberana, “Wireless-Optical Network Convergence: Enabling the 5G Architecture to Support Operational and End-User Services,” IEEE Comm. Mag., vol. 55, no. 10, pp. 184-192, 2017.
- [SNIM-3] “Linux KVM,” [Online]. Available: <https://www.linux-kvm.org>
- [SNIM-4] “Dell, VMware,” [Online]. Available: <https://www.vmware.com/>
- [SNIM-5] M. M. M., D. D. D. and K. D., “VXLAN, A Framework for Overlaying Virtualized Layer 2 Networks Over Layer 3 Networks,” Internet Engineering Task Force, 2011.
- [SNIM-6] M. Sridharan, K. Duda, I. Ganga, A. Greenberg, G. Lin and M. P., “NVGRE: network virtualization using generic routing encapsulation,” Internet Engineering Task Force, 2011.
- [SNIM-7] ETSI GS NFV MAN. "Networks Functions Virtualization (NFV); Management and Orchestration". Dec. 2014. V1.1.1.
- [SNIM-8] VNF and CNF, what's the difference? (no date) Red Hat - We make open source technologies for the enterprise. Available at: <https://www.redhat.com/en/topics/cloud-native-apps/vnf-and-cnf-whats-the-difference> (Accessed: December 2, 2022).
- [SNIM-9] Cncf (no date) CNCF/trailmap: 🗺️trailmap files from the CNCF/Landscape Repo, GitHub. Available at: <https://github.com/cncf/trailmap> (Accessed: December 2, 2022).
- [SNIM-10] “Cloud Native Network Functions Design, Architecture and Technology Landscape,” 2019, Accessed: Nov. 25, 2022. [Online]. Available: www.metaswitch.com
- [SNIM-11] D. Santos, R. Silva, D. Corujo, R. L. Aguiar and B. Parreira, "Follow the User: A Framework for Dynamically Placing Content Using 5G-Enablers," in IEEE Access, vol. 9, pp. 14688-14709, 2021, doi: 10.1109/ACCESS.2021.3051570.

Resource orchestration approaches

- [ROA-1] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network Slicing & Softwarization: A Survey on Principles, Enabling Technologies & Solutions,” in IEEE Communications Surveys & Tutorials, 3rd Quarter, Vol. 20, No. 3, Mar. 2018, pp. 2429-2453.
- [ROA-2] GSNFV, E. (2015). Network Functions Virtualisation (NFV); Ecosystem; Report on SDN Usage in NFV Architectural Framework V 1.1.1.
- [ROA-3] GSNFV, E. (2014). Network functions virtualisation (NFV): Architectural framework V1.2.1.
- [ROA-4] GSNFV, E. (2022). Network Functions Virtualisation (NFV) Release 3; Protocols and Data Models; NFV descriptors based on TOSCA specification v 3.6.1.
- [ROA-5] GSNFV, E. (2022). Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Multiple Administrative Domain Aspect Interfaces Specification V3.6.1.
- [ROA-6] GSNFV, E. (2022). Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on architecture options to support multiple administrative domains.
- [ROA-7] A. Laghrissi and T. Taleb, “A Survey on the Placement of Virtual Resources and Virtual Network Functions,” in IEEE Communications Surveys & Tutorials, Vol. 21, No. 2, 2nd Quarter 2019, pp. 1409 – 1434

- [ROA-8] T. Taleb, I. Afolabi, K. Samdanis and F. Z. Yousaf, “On Multi-domain Network Slicing Orchestration Architecture & Federated Resource Control,” in IEEE Network Magazine, Vol. 33, No. 5, Sep. 2019, pp. 242 - 252.
- [ROA-9] C. Benzaid and T. Taleb, “AI-driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions,” in IEEE Network Magazine, Vol. 34, No. 2, Mar. 2020, pp. 186-194
- [ROA-10] C. Benzaid and T. Taleb, “ZSM Security: Threat Surface and Best Practices,” in IEEE Network Magazine, Vol. 34, No. 3, Jun. 2020, pp. 124 – 133
- [ROA-11] C. Benzaid, T. Taleb, C.T. Phan, C. Tselios, and G. Tsolis, “Distributed AI-based Security for Massive Numbers of Network Slices in 5G & Beyond Mobile Systems,” in Proc. of 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Porto, Portugal, Jun. 2021.
- [ROA-12] A. Aissioui, A. Ksentini, A. Gueroui, and T. Taleb, “Toward Elastic Distributed SDN/NFV Controller for 5G Mobile Cloud Management Systems” in IEEE Access Magazine, DOI 10.1109/ACCESS.2015.2489930, Vol. 3, Nov. 2015
- [ROA-13] C. Benzaid and T. Taleb, “AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler?” in IEEE Network Magazine, Vol. 34, No. 6, Nov. 2020, pp. 140 - 147.
- [ROA-14] ETSI GS ZSM 002, “Zero-touch Network and Service Management (ZSM); Reference Architecture,” Aug. 2019
- [ROA-15] S. Kianpisheh and T. Taleb, “A Survey on In-network Computing: Programmable Data Plane And Technology Specific Applications,” to appear in IEEE COMST.
- [ROA-16] O. Abdul Wahab, Omar, A. Mourad, H. Otrok, and T. Taleb, “Federated Machine Learning: Survey, Multi-Level Classification, Desirable Criteria and Future Directions in Communication and Networking Systems,” in IEEE COMST, Vol. 23, No. 2, Secondquarter 2021, pp. 1342 – 1397

APIs, monitoring and communication services for the continuum

- [SO-1] Optimisation Concepts — PuLP 2.6.0 documentation, 2009.
- [SO-2] IBM Documentation, Mar. 2021.
- [SO-3] P. Buschmann, M. H. M. Shorim, M. Helm, A. Bröring, and G. Carle. Task allocation in industrial edge networks with particle swarm optimization and deep reinforcement learning. In 12th International Conference on the Internet of Things, IoT ’22, New York, NY, USA, 2022. Association for Computing Machinery.
- [SO-4] V. Cardellini, V. Grassi, F. Lo Presti, and M. Nardelli. Optimal operator placement for distributed stream processing applications. In Proceedings of the 10th ACM International Conference on Distributed and Event-based Systems, pages 69–80, Irvine California, June 2016. ACM.
- [SO-5] Y. Gao and Y. Wang. Multiple Workflows Offloading Based on Deep Reinforcement Learning in Mobile Edge Computing. In Y. Lai, T. Wang, M. Jiang, G. Xu, W. Liang, and A. Castiglione, editors, Algorithms and Architectures for Parallel Processing, Lecture Notes in Computer Science, pages 476–493, Cham, 2022. Springer International Publishing.
- [SO-6] M. Lesche. Framework, 2022.
- [SO-7] J. Seeger, A. Bröring, and G. Carle. Optimally Self-Healing IoT Choreographies, July 2019. arXiv:1907.04611 [cs].
- [SO-8] O. Skarlat and S. Schulte. FogFrame: a framework for IoT application execution in the fog. PeerJ Computer Science, July 2021.
- [SO-9] Q. You and B. Tang. Efficient task offloading using particle swarm optimization algorithm in edge computing for industrial internet of things. Journal of Cloud Computing, 10(1):41, July 2021.

- [SO-10] McDonnell, Tyler, Baishakhi Ray, und Miryung Kim. „An Empirical Study of API Stability and Adoption in the Android Ecosystem“. In 2013 IEEE International Conference on Software Maintenance, 70–79. Eindhoven, Netherlands: IEEE, 2013. <https://doi.org/10.1109/ICSM.2013.18>.
- [SO-11] Eilertsen, Anna Maria, und Anya Helene Bagge. „Exploring API: Client Co-Evolution“. In Proceedings of the 2nd International Workshop on API Usage and Evolution, 10–13. Gothenburg Sweden: ACM, 2018. <https://doi.org/10.1145/3194793.3194799>.
- [SO-12] Yu, Shuli, und C. Jason Woodard. „Innovation in the Programmable Web: Characterizing the Mashup Ecosystem“. In Service-Oriented Computing – ICSOC 2008 Workshops, herausgegeben von George Feuerlicht und Winfried Lamersdorf, 136–47. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2009. https://doi.org/10.1007/978-3-642-01247-1_13.
- [SO-13] Um, SungYong, und Youngjin Yoo. „The Co-Evolution of Digital Ecosystems“, o. J., 15.
- [SO-14] Andreo, Sebastien, und Jan Bosch. „API Management Challenges in Ecosystems“. In Software Business, herausgegeben von Sami Hyrnsalmi, Mari Suoranta, Anh Nguyen-Duc, Pasi Tyrväinen, und Pekka Abrahamsson, 86–93. Lecture Notes in Business Information Processing. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-33742-1_8.
- [SO-15] Yu, Shuli, und C. Jason Woodard. „Innovation in the Programmable Web: Characterizing the Mashup Ecosystem“. In Service-Oriented Computing – ICSOC 2008 Workshops, herausgegeben von George Feuerlicht und Winfried Lamersdorf, 136–47. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2009. https://doi.org/10.1007/978-3-642-01247-1_13.
- [SO-16] R. Kazman and H.-M. Chen, “The metropolis model and its implications for the engineering of software ecosystems,” in Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research – FoSER ’10. Santa Fe, New Mexico, USA: ACM Press, 2010, p. 187.
- [SO-17] S. Kolak, A. Afzal, C. Le Goues, M. Hilton, and C. S. Timperley, “It Takes a Village to Build a Robot: An Empirical Study of The ROS Ecosystem,” in 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME). Adelaide, Australia: IEEE, Sep. 2020, pp. 430–440.
- [SO-18] A. Alami, R. Pardo, M. Leavitt Cohn, and A. Wasowski, “Pull Request Governance In Open Source Communities,” IEEE Transactions on Software Engineering, pp. 1–1, 2021.
- [SO-19] J. S. Menzefricke, M. Frank, M. Drewel, and R. Dumitrescu, “Value-centered design of a digital service robotics platform,” Procedia CIRP, vol. 91, pp. 690–695, 2020.
- [SO-20] U. Naik und D. Shivalingaiah, „Comparative Study of Web 1.0, Web 2.0 and Web 3.0“, S. 9.
- [SO-21] B. Cyphers und C. Doctorow, „Privacy Without Monopoly: Data Protection and Interoperability“, S. 40.
- [SO-22] A. Hein, “Digital platform ecosystems,” p. 12, doi: doi.org/10.1007/s12525-019-00377-4.
- [SO-23] A. Marzinotto, M. Colledanchise, C. Smith and P. Ögren, “Towards a unified behavior trees framework for robot control,” 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 5420–5427, doi: [10.1109/ICRA.2014.6907656](https://doi.org/10.1109/ICRA.2014.6907656).
- [SO-24] R. Creemers, “14th Five-Year Plan for National Informatization,” Jan. 24, 2022. <https://digichina.stanford.edu/work/translation-14th-five-year-plan-for-national-informatization-dec-2021/>
- [SO-25] A. C. Bock and U. Frank, “Low-Code Platform,” Business & Information Systems. Engineering, vol. 63, no. 6, pp. 733–740, Dec. 2021.
- [SO-26] Alwen, J., Coretti, S., Dodis, Y. (2019). The Double Ratchet: Security Notions, Proofs, and Modularization for the Signal Protocol. In: Ishai, Y., Rijmen, V. (eds) Advances in Cryptology – EUROCRYPT 2019. EUROCRYPT 2019. Lecture Notes in Computer Science(), vol 11476. Springer, Cham. https://doi.org/10.1007/978-3-030-17653-2_5
- [SO-27] Aravindh Raman, Sagar Joglekar, Emiliano De Cristofaro, Nishanth Sastry, and Gareth Tyson. 2019. Challenges in the Decentralised Web: The Mastodon Case. In Proceedings of the Internet Measurement

Conference (IMC '19). Association for Computing Machinery, New York, NY, USA, 217–229. <https://doi.org/10.1145/3355369.3355572>

[SO-28] T. Raffin, T. Reichenstein, D. Klier, A. Kühn, J. Franke. Qualitative assessment of the impact of manufacturing-specific influences on Machine Learning Operations

[SO-29] L. Kidane, P. Townend, T. Metsch, E. Elmroth. When and How to Retrain Machine Learning-based Cloud Management Systems

[SO-30] R. M. Monarch. Human-in-the-Loop, Machine Learning

Data Orchestration Approaches

[DOA-1] Gartner, ‘Using Data Fabric Architecture to Modernize Data Integration’, *Gartner*. <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration> (accessed Oct. 23, 2022).

[DOA-2] Google, ‘Introducing the Knowledge Graph: things, not strings’, *Google*, May 16, 2012. <https://blog.google/products/search/introducing-knowledge-graph-things-not/> (accessed Oct. 25, 2022).

[DOA-3] S. Martin, B. Szekely, and D. Allemang, *The Rise of the Knowledge Graph*. 2021. Accessed: Oct. 24, 2022. [Online]. Available: <https://learning.oreilly.com/library/view/the-rise-of/9781098100407/>

[DOA-4] J. Barrasa, J. Webber, and A. Hodler, *Knowledge Graphs*. O’Reilly Media, Inc., 2021. Accessed: Oct. 20, 2022. [Online]. Available: <https://learning.oreilly.com/library/view/knowledge-graphs/9781098104863/>

[DOA-5] Delphix, ‘Data Virtualization’, *Delphix*. <https://www.delphix.com/glossary/what-is-data-virtualization> (accessed Oct. 24, 2022).

[DOA-6] IBM, ‘Data Fabric Solutions | IBM’. <https://www.ibm.com/data-fabric> (accessed Oct. 23, 2022).

[DOA-7] K2View, ‘Data Fabric Architecture | K2View’. <https://www.k2view.com/platform/data-fabric-architecture/> (accessed Oct. 25, 2022).

[DOA-8] Informatica, ‘What Is Data Fabric?’, *Informatica*. <https://www.informatica.com/resources/articles/data-fabric-the-transformative-next-step-in-data-management.html> (accessed Oct. 23, 2022).

[DOA-9] data.world, ‘The data catalog for Data Fabric’, *data.world*. <https://data.world/solutions/data-fabric/> (accessed Oct. 25, 2022).

[DOA-10] Stardog, ‘Use Cases - Data Fabric | Stardog’, *Stardog Union*. <https://www.stardog.com/use-cases/data-fabric/> (accessed Oct. 25, 2022).

[DOA-11] Talend, ‘Talend Data Fabric: The Complete Data Integration Platform | Talend’. <https://www.talend.com/products/data-fabric/> (accessed Oct. 25, 2022).

[DOA-12] Z. Dehghani, *Data Mesh: Delivering Data-Driven Value at Scale*. O’Reilly Media, 2022.

[DOA-13] ‘Augmented data management: Data fabric versus data mesh’, *Journey to AI Blog*, Apr. 27, 2022. <https://www.ibm.com/blogs/journey-to-ai/2022/04/augmented-data-management-data-fabric-versus-data-mesh/> (accessed Oct. 27, 2022).

Real time containers in the Industry

[RTC-1] M. Cinque, R. Della Corte, and R. Ruggiero, “Preventing timing failures in mixed-criticality clouds with dynamic real-time containers,” in 2021 17th European Dependable Computing Conference (EDCC), Sep. 2021, pp. 17–24. doi: 10.1109/EDCC53658.2021.00010.

[RTC-2] T. Cucinotta, L. Abeni, M. Marinoni, R. Mancini, and C. Vitucci, “Strong Temporal Isolation among Containers in OpenStack for NFV Services,” *IEEE Transactions on Cloud Computing*, pp. 1–1, 2021, doi: 10.1109/TCC.2021.3116183.

[RTC-3] S. Fiori, L. Abeni, and T. Cucinotta, “RT-Kubernetes - Containerized Real-Time Cloud Computing,” in ACM/SIGAPP Symposium on Applied Computing, Virtual Event. ACM, New York, NY, USA, Apr. 2022, p. 4. doi: 10.1145/3477314.3507216.

[RTC-4] S. Kuenzer, V. Badoiu, H. Lefeuvre, S. Santhanam, A. Jung, G. Gain, C. Soldani, C. Lupu, S. Teodorescu, C. Raducanu, C. Banu, L. Mathy, R. Deaconescu, C. Raiciu, and F. Huici, “Unikraft: fast, specialized unikernels the easy way,” in EuroSys ’21: Sixteenth European Conference on Computer Systems, Online Event, United Kingdom, April 26-28, 2021, A. Barbalace, P. Bhatotia, L. Alvisi, and C. Cadar, Eds. ACM, 2021, pp. 376–394. [Online]. Available: <https://doi.org/10.1145/3447786.3456248>

Edge-native approaches: cloud-native techniques applied along the computing continuum

[ENA-1] runc: CLI tool for spawning and running containers according to the OCI specification <https://github.com/opencontainers/runc>

[ENA-2] crun: a fast and lightweight fully featured OCI runtime <https://github.com/containers/crun>

[ENA- 3] Cri-o: lightweight container runtime for Kubernetes <https://cri-o.io/>

[ENA- 4] BalenaOS: run Docker containers on embedded devices <https://www.balena.io/os/>

[ENA- 5] Why Embedded Linux Needs a Container Manager Written in C <https://pantacor.com/blog/embedded-linux-need-container-manager/>

[ENA- 6] Pantavisor Linux: a framework for Containerized Embedded Linux <https://pantavisor.io/>

[ENA- 7] EVE-OS: edge Virtualization Engine <https://www.lfedge.org/projects/eve/>

[ENA- 8] K3s: lightweight Kubernetes <https://k3s.io/>

[ENA- 9] k3OS: the Kubernetes Operating System <https://k3os.io/>

[ENA- 10] MicroK8s: the lightweight Kubernetes <https://microk8s.io/>

[ENA- 11] KubeEdge: a Kubernetes native edge computing framework <https://kubedge.io/en/>

[ENA- 12] Cloud Native Computing Foundation projects Landscape <https://landscape.cncf.io/>

[ENA- 13] Sedna: an edge-cloud synergy AI project <https://sedna.readthedocs.io/en/latest/>

[ENA- 14] Open Horizon: containerized application deployment and lifecycle management, ML model synchronization to devices and Kubernetes clusters. <https://www.lfedge.org/projects/openhorizon/>

[ENA- 15] Baetyl: extend cloud computing, data and service seamlessly to edge devices <https://baetyl.io/en/>

[ENA- 16] Akri: a Kubernetes Resource Interface for the Edge <https://docs.akri.sh/>

[ENA- 17] OpenFaaS: serverless functions made simple <https://www.openfaas.com/>

[ENA- 18] Knative: serverless containers in Kubernetes environments <https://knative.dev/docs/>

[ENA- 19] CloudEvents: a specification for describing event data in a common way <https://cloudevents.io/>

[ENA- 20] AWS Greengrass features <https://aws.amazon.com/greengrass/features>

[ENA- 21] AWS Snowball Edge <https://docs.aws.amazon.com/snowball/latest/developer-guide/whatisedge.html>

[ENA- 22] Microsoft Azure IoT Edge <https://azure.microsoft.com/products/iot-edge/>

[ENA- 23] Microsoft Azure certified devices <https://www.microsoft.com/azure/partners/azure-certified-device>

[ENA- 24] Microsoft Azure Edge Stack <https://azure.microsoft.com/en-us/products/azure-stack/edge/>

[ENA- 25] Google Distributed Cloud <https://cloud.google.com/distributed-cloud>

- [ENA- 26] Kata Containers: open-source container runtime, building lightweight virtual machines that seamlessly plug into the containers ecosystem <https://katacontainers.io/>
- [ENA- 27] Madhavapeddy, A., & Scott, D. J. (2013). Unikernels: Rise of the Virtual Library Operating System: What if all the software layers in a virtual appliance were compiled within the same safe, high-level language framework? Queue, 11(11), 30-44.
- [ENA- 28] Simon Kienzler: Welcome To The Container Jungle: Docker vs. containerd vs. Nabla vs. Kata vs. Firecracker and more! Available online: <https://www.inovex.de/de/blog/containers-docker-containerd-nabla-kata-firecracker/>
- [ENA- 29] MirageOS: A programming framework for building type-safe, modular systems <https://mirage.io/>
- [ENA- [30] Nabla containers: a new approach to container isolation <https://nabla-containers.github.io/>
- [ENA- 31] WebAssembly: a binary instruction format for a stack-based virtual machine <https://webassembly.org/>
- [ENA- 32] WebAssembly System Interface: a modular system interface for WebAssembly <https://github.com/WebAssembly/WASI>
- [ENA- 33] Wasmtime: a fast and secure runtime for WebAssembly <https://wasmtime.dev/>
- [ENA- 34] WasmEdge: bring the cloud-native and serverless application paradigms to Edge Computing. <https://wasmedge.org/>
- [ENA- 35] Michael Irwing: introducing the Docker+Wasm Technical Preview. Available online: <https://www.docker.com/blog/docker-wasm-technical-preview/>

Self-* capabilities of heterogeneous nodes

- [SELF-1] W. Liu, "Research on cloud computing security problem and strategy," 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), 2012, pp. 1216-1219, doi: 10.1109/CECNet.2012.6202020
- [SELF-2] Mell, P. and Grance, T. (2011), The NIST Definition of Cloud Computing, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], <https://doi.org/10.6028/NIST.SP.800-145>
- [SELF-3] W. Yu et al., "A Survey on the Edge Computing for the Internet of Things," in IEEE Access, vol. 6, pp. 6900-6919, 2018, doi: 10.1109/ACCESS.2017.2778504
- [SELF-4] M. Xu and R. Buyya, "Managing renewable energy and carbon footprint in multi-cloud computing environments", in Journal of Parallel and Distributed Computing, vol. 135, pp. 191-202, 2020, doi: 10.1016/j.jpdc.2019.09.015
- [SELF-5] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198
- [SELF-6] Edge Computing Consortium & Alliance of Industrial Internet. Edge Computing Reference Architecture 2.0. Technical Report, Edge Computing Consortium, 2017. Available online: <http://en.eccconsortium.net/Uploads/file/20180328/1522232376480704.pdf>
- [SELF-7] J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu, "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues," in IEEE Access, vol. 6, pp. 18209-18237, 2018, doi: 10.1109/ACCESS.2018.2820162
- [SELF-8] M. A. Razzaque, M. Milojevic-Jevric, A. Palade and S. Clarke, "Middleware for Internet of Things: A Survey," in IEEE Internet of Things Journal, vol. 3, no. 1, pp. 70-95, Feb. 2016, doi: 10.1109/JIOT.2015.2498900

- [SELF-9] A. Xiao, Z. Lu, X. Du, J. Wu and P. C. K. Hung, "ORHRC: Optimized Recommendations of Heterogeneous Resource Configurations in Cloud-Fog Orchestrated Computing Environments," 2020 IEEE International Conference on Web Services (ICWS), 2020, pp. 404-412, doi: 10.1109/ICWS49710.2020.00059
- [SELF-10] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," in *Computer*, vol. 36, no. 1, pp. 41-50, Jan. 2003, doi: 10.1109/MC.2003.1160055
- [SELF-11] A. Berns and S. Ghosh, "Dissecting Self-* Properties," 2009 Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems, 2009, pp. 10-19, doi: 10.1109/SASO.2009.25
- [SELF-12] R. Sterritt and M. Hinchey, "SPAACE IV: Self-Properties for an Autonomous & Autonomic Computing Environment – Part IV A Newish Hope," 2010 Seventh IEEE International Conference and Workshops on Engineering of Autonomic and Autonomous Systems, 2010, pp. 119-125, doi: 10.1109/EASe.2010.29
- [SELF-13] D. Rosendo, A. Costan, P. Valduriez and G. Antoniu, "Distributed intelligence on the Edge-to-Cloud Continuum: A systematic literature review", in *Journal of Parallel and Distributed Computing*, vol. 166, pp. 71-94, 2022, doi: 10.1016/j.jpdc.2022.04.004
- [SELF-14] P. Hu, S. Dhelim, H. Ning and T. Qiu, "Survey on fog computing: architecture, key technologies, applications and open issues", in *Journal of Network and Computer Applications*, vol. 98, pp. 27-42, 2017, doi: 10.1016/j.jnca.2017.09.002
- [SELF-15] Y. Yu, "Mobile edge computing towards 5G: Vision, recent progress, and open challenges," in *China Communications*, vol. 13, no. Supplement2, pp. 89-99, 2016, doi: 10.1109/CC.2016.7833463
- [SELF-16] Xiaopei Wu, Robert Dunne, Qingyang Zhang, and Weisong Shi. 2017. Edge computing enabled smart firefighting: opportunities and challenges. In *Proceedings of the fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb '17)*. Association for Computing Machinery, New York, NY, USA, Article 11, 1–6. <https://doi.org/10.1145/3132465.3132475>
- [SELF-17] M. Antonini, M. Pincheira, M. Vecchio and F. Antonelli, "Tiny-MLOps: a framework for orchestrating ML applications at the far edge of IoT systems," 2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2022, pp. 1-8, doi: 10.1109/EAIS51927.2022.9787703
- [SELF-18] M. Götzinger et al., "RoSA: A Framework for Modeling Self-Awareness in Cyber-Physical Systems," in *IEEE Access*, vol. 8, pp. 141373-141394, 2020, doi: 10.1109/ACCESS.2020.3012824
- [SELF-19] A. Diaconescu, B. Porter, R. Rodrigues and E. Pournaras, "Hierarchical Self-Awareness and Authority for Scalable Self-Integrating Systems," 2018 IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W), 2018, pp. 168-175, doi: 10.1109/FAS-W.2018.00043
- [SELF-20] L. Esterle and J. N. Brown, "Levels of Networked Self-Awareness," 2018 IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W), 2018, pp. 237-238, doi: 10.1109/FAS-W.2018.00054
- [SELF-21] P. R. Lewis et al., "Architectural Aspects of Self-Aware and Self-Expressive Computing Systems: From Psychology to Engineering," in *Computer*, vol. 48, no. 8, pp. 62-70, Aug. 2015, doi: 10.1109/MC.2015.235
- [SELF-22] A. Anzanpour et al., "Self-awareness in remote health monitoring systems using wearable electronics," *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2017, 2017, pp. 1056-1061, doi: 10.23919/DATE.2017.7927146
- [SELF-23] R. Andrade and J. Torres, "Self-Awareness as an enabler of Cognitive Security," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 701-708, doi: 10.1109/IEMCON.2018.8614798
- [SELF-24] D. Sinreich, "An architectural blueprint for autonomic computing," IBM Corp., Armonk, NY, USA, White paper, 2006. [Online]. Available: <https://www-03.ibm.com/autonomic/pdfs/AC%20Blueprint%20White%20Paper%20V7.pdf>

- [SELF-25] A. Elhabbash, R. Bahsoon, P. Tino, P. R. Lewis and Y. Elkhatab, "Attaining Meta-self-awareness through Assessment of Quality-of-Knowledge," 2021 IEEE International Conference on Web Services (ICWS), 2021, pp. 712-723, doi: 10.1109/ICWS53863.2021.00099
- [SELF-26] C. S. Regazzoni, L. Marcenaro, D. Campo and B. Rinner, "Multisensorial Generative and Descriptive Self-Awareness Models for Autonomous Systems," in *Proceedings of the IEEE*, vol. 108, no. 7, pp. 987-1010, July 2020, doi: 10.1109/JPROC.2020.2986602
- [SELF-27] N. Zhang, R. Bahsoon and G. Theodoropoulos, "Towards Engineering Cognitive Digital Twins with Self-Awareness," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 3891-3891, doi: 10.1109/SMC42975.2020.9283357
- [SELF-28] I. M. Delamer and J. L. M. Lastra, "Self-orchestration and choreography: towards architecture-agnostic manufacturing systems," 20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA'06), 2006, pp. 5 pp.-, doi: 10.1109/AINA.2006.301
- [SELF-29] Z. Wen, R. Yang, P. Garraghan, T. Lin, J. Xu and M. Rovatsos, "Fog Orchestration for Internet of Things Services," in *IEEE Internet Computing*, vol. 21, no. 2, pp. 16-24, Mar.-Apr. 2017, doi: 10.1109/MIC.2017.36
- [SELF-30] K. Khebbab, N. Hameurlain and F. Belala, "A Maude-Based rewriting approach to model and verify Cloud/Fog self-adaptation and orchestration", in *Journal of Systems Architecture*, vol. 110, 2020, ISSN 1383-7621, doi: 10.1016/j.sysarc.2020.101821
- [SELF-31] M. Ruta, F. Scioscia, G. Loseto and E. D. Sciascio, "A Semantic-Enabled Social Network of Devices for Building Automation," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3379-3388, Dec. 2017, doi: 10.1109/TII.2017.2697907
- [SELF-32] D. Schulz, "Intent-based automation networks: Toward a common reference model for the self-orchestration of industrial intranets," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 2016, pp. 4657-4664, doi: 10.1109/IECON.2016.7792959
- [SELF-33] F. M. Discenzo, P. J. Unsworth, K. A. Loparo and H. O. Marcy, "Self-diagnosing intelligent motors: a key enabler for next generation manufacturing systems," *IEE Colloquium on Intelligent and Self-Validating Sensors* (Ref. No. 1999/160), 1999, pp. 3/1-3/4, doi: 10.1049/ic:19990763
- [SELF-34] Van-Trinh Hoang, Nathalie Julien and Pascal Berruet. On-line self-diagnosis based on power measurement for a wireless sensor node. First IEEE Workshop on Highly-Reliable Power-Efficient Embedded Designs, Feb 2013, Shenzhen, China. [Online]. Available: <http://hal.univ-brest.fr/hal-00782758v2>
- [SELF-35] V. Volotka, "Methods of self-diagnosing in telecommunication networks," 2015 Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), 2015, pp. 131-134, doi: 10.1109/INFOCOMMST.2015.7357292
- [SELF-36] S. A. Raheem, M. Prabhakar and C. Venugopal, "Comb needle model for data aggregation using self-diagnose cluster in WSN," 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), 2017, pp. 390-394, doi: 10.1109/SmartTechCon.2017.8358402
- [SELF-37] S. Harte, A. Rahman and K. M. Razeed, "Fault tolerance in sensor networks using self-diagnosing sensor nodes," *The IEE International Workshop on Intelligent Environments*, 2005 (Ref. No. 2005/11059), 2005, pp. 7-12, doi: 10.1049/ic:20050211
- [SELF-38] M. Elhadeif, A. Boukerche and H. Elkadiki, "Self-Diagnosing Wireless Mesh and Ad-Hoc Networks using an Adaptable Comparison-Based Approach," *The Second International Conference on Availability, Reliability and Security (ARES'07)*, 2007, pp. 983-990, doi: 10.1109/ARES.2007.140
- [SELF-39] H. -Y. Cheng and L. -W. Tsai, "Balancing robustness and information abundance via self-diagnosing in traffic surveillance video analysis," 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2015, pp. 1-6, doi: 10.1109/AVSS.2015.7301726
- [SELF-40] D. Xikun, W. Huiqiang and L. Hongwu, "A Comprehensive Monitor Model for Self-Healing Systems," 2010 International Conference on Multimedia Information Networking and Security, 2010, pp. 751-756, doi: 10.1109/MINES.2010.159

- [SELF-41] K. Khalil, O. Eldash, A. Kumar and M. Bayoumi, "Self-Healing Approach for Hardware Neural Network Architecture," 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), 2019, pp. 622-625, doi: 10.1109/MWSCAS.2019.8885235
- [SELF-42] L. YANG, F. XIAO, H. CHEN, Y. LAI and Y. CHOLLOT, "The Experiences of Decentralized Self-Healing Grid," 2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP), 2019, pp. 1864-1867, doi: 10.1109/APAP47170.2019.9225046
- [SELF-43] W. Liu, T. Kang, W. Cheng and F. Zhao, "The modelling of self-healing control system for distribution network based on UML," 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2015, pp. 1435-1439, doi: 10.1109/DRPT.2015.7432458
- [SELF-44] Z. Liu, C. Gui and C. Ma, "Design and Verification of Integrated Ship Monitoring Network with High Reliability and Zero-time Self-healing," 2019 Chinese Control And Decision Conference (CCDC), 2019, pp. 2348-2351, doi: 10.1109/CCDC.2019.8832456
- [SELF-45] J. Hou, "A method of distribution network reconstruction based on self-healing technology," 2021 China International Conference on Electricity Distribution (CICED), 2021, pp. 784-788, doi: 10.1109/CICED50259.2021.9556720
- [SELF-46] N. R. Herbst, S. Kounev and R. Reussner, "Elasticity in cloud computing: What it is, and what it is not.", in 10th International Conference on Autonomic Computing (ICAC), pp 23–27, 2013.
- [SELF-47] J. Herrera and G. Moltó, "Toward Bio-Inspired Auto-Scaling Algorithms: An Elasticity Approach for Container Orchestration Platforms," in IEEE Access, vol. 8, pp. 52139-52150, 2020, doi: 10.1109/ACCESS.2020.2980852
- [SELF-48] A. Mehmood, T. A. Khan, J. J. Diaz Rivera and W. -C. Song, "Dynamic Auto-scaling of VNFs based on Task Execution Patterns," 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2019, pp. 1-4, doi: 10.23919/APNOMS.2019.8892836
- [SELF-49] A. Y. Nikraves, S. A. Ajila and C. -H. Lung, "Towards an Autonomic Auto-scaling Prediction System for Cloud Resource Provisioning," 2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, 2015, pp. 35-45, doi: 10.1109/SEAMS.2015.22
- [SELF-50] E. Casalicchio and V. Perciballi, "Auto-Scaling of Containers: The Impact of Relative and Absolute Metrics," 2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W), 2017, pp. 207-214, doi: 10.1109/FAS-W.2017.149
- [SELF-51] S. Chattopadhyay, S. Chatterjee, S. Nandi and S. Chakraborty, "Aloe: An Elastic Auto-Scaled and Self-stabilized Orchestration Framework for IoT Applications," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, 2019, pp. 802-810, doi: 10.1109/INFOCOM.2019.8737656
- [SELF-52] Parashar, M., Hariri, S. (2005). Autonomic Computing: An Overview. In: Banâtre, JP., Fradet, P., Giavitto, JL., Michel, O. (eds) Unconventional Programming Paradigms. UPP 2004. Lecture Notes in Computer Science, vol 3566. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11527800_20
- [SELF-53] Guangxiang Yang and Hua Liang, "Self configuration of 4G network terminals," 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), 2010, pp. 80-83, doi: 10.1109/CAR.2010.5456774
- [SELF-54] J. Z. Wang and M. Vanninen, "Self-Configuration Protocols for Small-Scale P2P Networks," 2006 IEEE/IFIP Network Operations and Management Symposium NOMS 2006, 2006, pp. 1-4, doi: 10.1109/NOMS.2006.1687675
- [SELF-55] L. Mombello, N. Calarco and F. P. Quintián, "System-on-Chip Implementation of a Self-Configuration System for a Programmable Photodetector ASIC," 2020 Argentine Conference on Electronics (CAE), 2020, pp. 99-103, doi: 10.1109/CAE48787.2020.9046361
- [SELF-56] E. Guan, J. Liu and Y. Zhao, "Self-configuration strategy design for unit-compressible modular robotic system," CSAA/IET International Conference on Aircraft Utility Systems (AUS 2020), 2020, pp. 232-237, doi: 10.1049/icp.2021.0150

- [SELF-57] G. ABDELLAOUI, H. MEGNAFI and F. T. BENDIMERAD, "A novel model using Reo for IoT self-configuration systems," 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP), 2020, pp. 1-5, doi: 10.1109/CCSSP49278.2020.9151679
- [SELF-58] J. Yao, Q. Lu and Z. Qi, "Automated Resource Sharing for Virtualized GPU with Self-Configuration," 2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS), 2017, pp. 250-252, doi: 10.1109/SRDS.2017.35
- [SELF-59] N. S. Bülül, D. Ergenç and M. Fischer, "SDN-based Self-Configuration for Time-Sensitive IoT Networks," 2021 IEEE 46th Conference on Local Computer Networks (LCN), 2021, pp. 73-80, doi: 10.1109/LCN52139.2021.9524979
- [SELF-60] M. R. Nami and K. Bertels, "A Survey of Autonomic Computing Systems," Third International Conference on Autonomic and Autonomous Systems (ICAS'07), 2007, pp. 26-26, doi: 10.1109/CONIELECOMP.2007.48
- [SELF-61] R. Zheng, M. Zhang, Q. Wu, G. Li and W. Wei, "A Service Self-Optimization Algorithm based on Autonomic Computing," 2009 IEEE International Conference on Granular Computing, 2009, pp. 805-808, doi: 10.1109/GRC.2009.5255010
- [SELF-62] I. Shaya, M. Ergen, A. Azizan, M. Ismail and Y. I. Daradkeh, "Individualistic Dynamic Handover Parameter Self-Optimization Algorithm for 5G Networks Based on Automatic Weight Function," in IEEE Access, vol. 8, pp. 214392-214412, 2020, doi: 10.1109/ACCESS.2020.3037048
- [SELF-63] J. Sánchez-González, J. Pérez-Romero and O. Sallent, "A Rule-Based Solution Search Methodology for Self-Optimization in Cellular Networks," in IEEE Communications Letters, vol. 18, no. 12, pp. 2189-2192, Dec. 2014, doi: 10.1109/LCOMM.2014.2363670
- [SELF-64] Trumler, W., Thiemann, T., Ungerer, T. (2006). An Artificial Hormone System for Self-organization of Networked Nodes. In: Pan, Y., Rammig, F.J., Schmeck, H., Solar, M. (eds) Biologically Inspired Cooperative Computing. BICC 2006. IFIP International Federation for Information Processing, vol 216. Springer, Boston, MA. doi: 10.1007/978-0-387-34733-2_9
- [SELF-65] E. Meshkova et al., "Designing a Self-Optimization System for Cognitive Wireless Home Networks," in IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 684-702, Dec. 2017, doi: 10.1109/TCCN.2017.2755010
- [SELF-66] L. Wang, J. Liu, Q. Wu and X. Wang, "Ship Course Control Based on BSO-PID Online Self-Optimization Algorithm," 2019 5th International Conference on Transportation Information and Safety (ICTIS), 2019, pp. 1405-1411, doi: 10.1109/ICTIS.2019.8883803
- [SELF-67] Cheng, B.H.C. et al. (2009). Software Engineering for Self-Adaptive Systems: A Research Roadmap. In: Cheng, B.H.C., de Lemos, R., Giese, H., Inverardi, P., Magee, J. (eds) Software Engineering for Self-Adaptive Systems. Lecture Notes in Computer Science, vol 5525. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-02161-9_1
- [SELF-68] C. Krupitzer, F. Roth, S. VanSyckel, G. Schiele and C. Becker, "A survey on engineering approaches for self-adaptive systems", in Pervasive and Mobile Computing, vol. 17, part B, pp. 184-206, 2015, doi: 10.1016/j.pmcj.2014.09.009
- [SELF-69] A. Amiri, U. Zdun, A. van Hoorn and S. Dustdar, "Automatic Adaptation of Reliability and Performance Trade-Offs in Service- and Cloud-Based Dynamic Routing Architectures," 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS), 2021, pp. 434-445, doi: 10.1109/QRS54544.2021.00055
- [SELF-70] M. Chen, Y. Wang, P. Li and H. Fu, "Research on an improved PSO algorithm with dual self-adaptation and dual variation," 2022 IEEE International Conference on Mechatronics and Automation (ICMA), 2022, pp. 646-650, doi: 10.1109/ICMA54519.2022.9856223
- [SELF-71] S. Zhang, M. Zhang, L. Ni and P. Liu, "A Multi-Level Self-Adaptation Approach For Microservice Systems," 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 498-502, doi: 10.1109/ICCCBDA.2019.8725647

- [SELF-72] V. Nallur and R. Bahsoon, "A decentralized self-adaptation mechanism for service-based applications in the cloud," in IEEE Transactions on Software Engineering, vol. 39, no. 5, pp. 591-612, May 2013, doi: 10.1109/TSE.2012.53
- [SELF-73] L. Ardito, "Energy aware self-adaptation in mobile systems," 2013 35th International Conference on Software Engineering (ICSE), 2013, pp. 1435-1437, doi: 10.1109/ICSE.2013.6606736
- [SELF-74] Y. Yuan, W. Zhang and X. Zhang, "A Context-Aware Self-Adaptation Approach for Web Service Composition," 2018 3rd International Conference on Information Systems Engineering (ICISE), 2018, pp. 33-38, doi: 10.1109/ICISE.2018.00014
- [SELF-75] S. R. Boyapati and C. Szabo, "Self-adaptation in Microservice Architectures: A Case Study," 2022 26th International Conference on Engineering of Complex Computer Systems (ICECCS), 2022, pp. 42-51, doi: 10.1109/ICECCS54210.2022.00014
- [SELF-76] S. Poslad, "Autonomous systems and Artificial Life", in Ubiquitous Computing Smart Devices, Smart Environments and Smart Interaction, 2009.
- [SELF-77] M. Dongzhi, Z. Jiangbin, Y. Xinping and Z. Tao, "Development of ME-GI dual-fuel engine fault diagnosis expert system based on self-learning ontology," 2014 Prognostics and System Health Management Conference (PHM-2014 Hunan), 2014, pp. 125-130, doi: 10.1109/PHM.2014.6988147
- [SELF-78] J. Zhang, X. Wen and L. Zeng, "Research of parameter self-learning fuzzy control strategy in motor control system for electric vehicles," 2009 International Conference on Electrical Machines and Systems, 2009, pp. 1-5, doi: 10.1109/ICEMS.2009.5382676
- [SELF-79] L. Wen-Bin, "A Self-learning Algorithm for Space Environment Temperature Control," 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control, 2012, pp. 1649-1652, doi: 10.1109/IMCCC.2012.383
- [SELF-80] N. Jamshidpour, E. H. Aria, A. Safari and S. Homayouni, "Adaptive Self-Learned Active Learning Framework for Hyperspectral Classification," 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2019, pp. 1-5, doi: 10.1109/WHISPERS.2019.8921298
- [SELF-81] Zhuang Ji-Hui, Xie Hui and Yan Ying, "Remote self-learning of driving cycle for electric vehicle demonstrating area," 2008 IEEE Vehicle Power and Propulsion Conference, 2008, pp. 1-4, doi: 10.1109/VPPC.2008.4677700
- [SELF-82] Zhongren Chen and Yejun He, "A smart power saver based on composite switch and self-learning fuzzy control for drinking water dispenser," 2016 IEEE International Conference on Power and Renewable Energy (ICPRE), 2016, pp. 275-278, doi: 10.1109/ICPRE.2016.7871215
- [SELF-83] P. Abeyasinghe and T. Bandara, "A novel self-learning approach to overcome incompatibility on TripAdvisor reviews", in Data Science and Management, vol. 5, issue 1, pp. 1-10, 2022, doi: 10.1016/j.dsm.2022.02.001

Data syntactic and semantic interoperability in the continuum

- [DIC-1] P. Wegner, "Interoperability," ACM Computing Surveys, vol. 28, no. 1, pp. 285-287, March 1996.
- [DIC-2] W. Wang, A. Tolk and W. Wang, "The levels of conceptual interoperability model: applying systems engineering principles to M&S," in Proceedings of the 2009 Spring Simulation Multiconference, 2009.
- [DIC-3] European Commission, "European Interoperability Framework for Pan-European e-Government Services," 2008, p. 79.
- [DIC-4] H. van der Veer and A. Wiles, "Achieving Technical Interoperability – the ETSI Approach," 2008.
- [DIC-5] "Extensible Markup Language (XML)," [Online]. Available: <https://www.w3.org/XML/>.

- [DIC-6] “ECMA-404: The JSON data interchange syntax, 2nd edition,” December 2017. [Online]. Available: <https://www.ecma-international.org/publications-and-standards/standards/ecma-404/>.
- [DIC-7] “RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation,” [Online]. Available: <https://www.w3.org/TR/rdf11-concepts/>.
- [DIC-8] “JSON-LD 1.1: A JSON-based Serialization for Linked Data. W3C Recommendation,” [Online]. Available: <https://www.w3.org/TR/json-ld11/>.
- [DIC-9] F. Galiegue, K. Zyp and G. Court, “JSON Schema: core definitions and terminology,” Internet Engineering Task Force (IETF), 2013.
- [DIC-10] “OWL 2 Web Ontology Language – Document Overview (Second Edition). W3C Recommendation,” [Online]. Available: <https://www.w3.org/TR/owl-overview/>.
- [DIC-11] “RDF Schema 1.1. W3C Recommendation,” [Online]. Available: <https://www.w3.org/TR/rdf-schema/>.
- [DIC-12] P. Chen, “The Entity-Relationship Model – Toward a Unified View of Data,” ACM Transactions on Database Systems, vol. 1, no. 1, pp. 9-36, 1976.
- [DIC-13] “Unified Modeling Language,” [Online]. Available: <http://www.omg.org/spec/UML/>.
- [DIC-14] “Semantic Web and Linked Data: Ontologies and Frameworks,” [Online]. Available: https://guides.library.ucla.edu/semantic-web/semantic_web_ontologies.
- [DIC-15] “Ontomalizer,” [Online]. Available: <https://github.com/srdc/ontmalizer>.
- [DIC-16] “ReDeFer,” [Online]. Available: <https://rhizomik.net/redefer>.
- [DIC-17] M. Ganzha, M. Paprzycki, W. Pawłowski, P. Szmeja and K. Wasielewska, “Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective,” Journal of Network and Computer Applications, vol. 81, pp. 111-124, 2017.
- [DIC-18] “Generic Ontology for IoT Platforms,” 8 March 2018. [Online]. Available: <https://inter-iot.github.io/ontology/>.
- [DIC-19] A. Guinard, “A survey of Top-Level Ontologies,” Digital Twin Hub, 2021.
- [DIC-20] S. Auer, J. Lehmann and S. Hellmann, “LinkedGeoData: Adding a Spatial Dimension to the Web of Data,” in The Semantic Web – ISWC 2009, 2009.
- [DIC-21] “geoSPARQL,” [Online]. Available: <https://www.opengeospatial.org/standards/geosparql>.
- [DIC-22] “Basic Geo (WGS84 lat/long) Vocabulary,” [Online]. Available: <https://www.w3.org/2003/01/geo>.
- [DIC-23] “Library for Quantity Kinds and Units: schema, based on QUDV model OMG SysML(TM), Version 1.2,” [Online]. Available: <https://www.w3.org/2005/Incubator/ssn/ssnx/qu/>.
- [DIC-24] “OM - Ontology of units of Measure,” [Online]. Available: <https://github.com/HajoRijgersberg/OM>.
- [DIC-25] “SWEET Ontologies,” [Online]. Available: <https://github.com/ESIPFed/sweet>.
- [DIC-26] “Time Ontology in OWL,” 26 March 2020. [Online]. Available: <https://www.w3.org/TR/owl-time/>.
- [DIC-27] “PROV-O: The PROV Ontology,” 30 April 2013. [Online]. Available: <https://www.w3.org/TR/prov-o>.
- [DIC-28] T. Schneider and M. Šimkus, “Ontologies and Data Management: A Brief Survey,” KI – Künstliche Intelligenz, vol. 34, pp. 329-353, 2020.
- [DIC-29] “Apache Kafka,” [Online]. Available: <https://kafka.apache.org/>.
- [DIC-30] “Apache Storm,” [Online]. Available: <https://storm.apache.org/>.

- [DIC-31] “Apache Flink,” [Online]. Available: <https://flink.apache.org/>.
- [DIC-32] M. Ganzha, M. Paprzycki, W. Pawłowski, B. Solarz-Niesłuchowski, P. Szmeja and K. Wasielewska, “Semantic Interoperability,” in *Interoperability of Heterogeneous IoT Platforms. A Layered Approach*, Springer International Publishing, 2021, pp. 133-165.
- [DIC-33] P. Sowiński, K. Wasielewska-Michniewska, M. Ganzha, W. Pawłowski, P. Szmeja and M. Paprzycki, “Efficient RDF Streaming for the Edge-Cloud Continuum,” in *IEEE 8th World Forum on Internet of Things*, Yokohama, Japan, 2022.

Data sovereignty, governance and lineage policies

- [DSGP-1] E. Eryurek, U. Gilad, V. Lakshmanan, A. Kibunguchy, J. Ashdown, and an O. M. C. Safari, *Data Governance: The Definitive Guide*. O'Reilly Media, Incorporated, 2021. [Online]. Available: <https://books.google.es/books?id=Z6XfzQEACAAJ>
- [DSGP-2] J. Reis and M. Housley, *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*. O'Reilly Media, Incorporated, 2022. [Online]. Available: <https://books.google.es/books?id=Z\TFzgEACAAJ>
- [DSGP-3] ‘Home - Egeria’. <https://egeria-project.org/> (accessed Oct. 27, 2022).
- [DSGP-4] ‘Apache Atlas – Data Governance and Metadata framework for Hadoop’. <https://atlas.apache.org/#/> (accessed Oct. 27, 2022).
- [DSGP-5] ‘Amundsen, the leading open source data catalog’. <https://www.amundsen.io/> (accessed Oct. 27, 2022).
- [DSGP-6] ‘A Metadata Platform for the Modern Data Stack | DataHub’. <https://datahubproject.io/> (accessed Oct. 27, 2022).

Advanced AI management approaches

- [AI-1] Q. Li et al., “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection,” in *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2021.3124599
- [AI-2] Sarker, I.H. “Machine Learning: Algorithms, Real-World Applications and Research Directions”. *SN COMPUT. SCI.* 2, 160 (2021), doi: 10.1007/s42979-021-00592-x
- [AI-3] K. Bonawitz et al., “Towards Federated Learning at Scale: System Design.” 2019 [Online]. Available: <http://arxiv.org/abs/1902.01046>
- [AI-4] ODSC – Open Data Science, “What is Federated Learning?” 2020 [Online]. Available: <https://medium.com/@ODSC/what-is-federated-learning-99c7fc9bc4f5>
- [AI-5] McMahan, H. B. et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data”, *AISTATS* (2017)
- [AI-6] Konečný, Jakub et al. “Federated Learning: Strategies for Improving Communication Efficiency”, *ArXiv abs/1610.05492* (2016): n. pag.
- [AI-7] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu and C. Guan, "Federated Transfer Learning for EEG Signal Classification," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), (2020), pp. 3040-3045, doi: 10.1109/EMBC44109.2020.9175344
- [AI-8] Geun Hyeong Lee et al., “Reliability and Performance Assessment of Federated Learning on Clinical Benchmark Data” (2020)
- [AI-9] Hard, Andrew et al. “Training Keyword Spotting Models on Non-IID Data with Federated Learning”, *ArXiv abs/2005.10406* (2020): n. pag.
- [AI-10] Hiessl, Thomas et al. “Industrial Federated Learning - Requirements and System Design”, *ArXiv abs/2005.06850* (2020): n. pag.

- [AI-11] Fan, Chenyou and Ping Liu. “Federated Generative Adversarial Learning”, *PRCV* (2020)
- [AI-12] Kholod, Ivan I. et al. “Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis”, *Sensors* (Basel, Switzerland) 21 (2021): n. pag.
- [AI-13] Phillips, P. , Hahn, C. , Fontana, P. , Yates, A. , Greene, K. , Broniatowski, D. and Przybocki, M. (2021), “Four Principles of Explainable Artificial Intelligence”, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, <https://doi.org/10.6028/NIST.IR.8312> 2022 [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=933399
- [AI-14] Gianluca De Lucia, Marco Lapegna, Diego Romano, “Towards explainable AI for hyperspectral image classification in Edge Computing environments”, *Computers and Electrical Engineering*, Volume 103, (2022), 108381, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2022.108381>
- [AI-15] Evchenko, Mikhail & Vanschoren, Joaquin & Hoos, Holger & Schoenauer, Marc & Sebag, Michèle, “Frugal Machine Learning”, arXiv:2111.03731 (2021)
- [AI-16] Deepdetect, “Frugal models: strategies for deep models with small data” 2021 [Online]. Available: <https://www.deepdetect.com/blog/22-frugal-models/>
- [AI-17] Bryce Murray, “The most important types of XAI” 2022 [Online]. Available: <https://towardsdatascience.com/the-most-important-types-of-xai-72f5beb9e77e#:~:text=This%20post%20discusses%20three%20critical,transparent%20AI%2C%20and%20in,teractable%20AI>
- [AI-18] Rezaei, Mahdi and Mahsa Shahidi, “Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review.” *Intelligence-Based Medicine* 3 (2020): 100005 – 100005
- [AI-19] Song, Yisheng, et al., "A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities", arXiv preprint arXiv:2205.06743 (2022)
- [AI-20] Weiss, K., Khoshgoftaar, T.M. & Wang, “D. A survey of transfer learning”, *J Big Data* 3, 9 (2016), <https://doi.org/10.1186/s40537-016-0043-6>
- [AI-21] R. Kaur, R. Kumar and M. Gupta, "Review on Transfer Learning for Convolutional Neural Network", 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), (2021), pp. 922-926, doi: 10.1109/ICAC3N53548.2021.9725474
- [AI-22] Dipanjan (DJ) Sarkar, “A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning”, 2018 [Online]. Available: <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>
- [AI-23] Wang, Yaqing & Yao, Quanming & Kwok, James & Ni, Lionel, “Generalizing from a Few Examples: A Survey on Few-shot Learning”, *ACM Computing Surveys*. 53, (2020), pp. 1-34, doi: 10.1145/3386252
- [AI-24] Dushkin, Roman & Andronov, Mikhail, “The Hybrid Design for Artificial Intelligence Systems”, (2021), doi: 10.1007/978-3-030-55180-3_13
- [AI-25] N. O’ Mahony, Sean Campbell, Anderson Carvalho, L. Krpalkova, Gustavo Velasco Hernandez, Suman Harapanahalli, D. Riordan, J. Walsh, “One-Shot Learning for Custom Identification Tasks; A Review”, *Procedia Manufacturing*, Volume 38, (2019), pp. 186-193, ISSN 2351-9789, doi: 10.1016/j.promfg.2020.01.025
- [AI-26] Yu, H., Mineyev, I., Varshney, L.R., & Evans, J.A., “Learning from One and Only One Shot”, ArXiv, abs/2201.08815 (2022)
- [AI-27] Dataconomy, “Active Learning overcomes the ML training challenges”, 2022 [Online]. Available: <https://dataconomy.com/2022/09/active-learning-machine-learning/>

Security, integrity, trust, privacy and policy enforcement in the computing continuum

- [SCC-1] D. Liu, Z. Yan, W. Ding y M. Atiquzzaman, «A Survey on Secure Data Analytics in Edge Computing,» IEEE Internet of Things Journal, vol. 6, pp. 4946-4967, 2019.
- [SCC-2] P. J. Sun, «Privacy Protection and Data Security in Cloud Computing: A Survey, Challenges, and Solutions,» IEEE Access, vol. 7, pp. 147420-147452, 2019.
- [SCC-3] A. Alwarafy, K. A. Al-Thelaya, M. Abdallah, J. Schneider y M. Hamdi, «A Survey on Security and Privacy Issues in Edge-Computing-Assisted Internet of Things,» IEEE Internet of Things Journal, vol. 8, pp. 4004-4022, 2020.
- [SCC-4] N. Chunchun , C. Li Shan, G. Prosanta y M. Geyong, «Data anonymization evaluation for big data and IoT environment,» Information Sciences, vol. 605, pp. 381-392, 2022.
- [SCC-5] P. Silva, E. Monteiro y P. Simões, «Privacy in the Cloud: A Survey of Existing Solutions and Research Challenges,» IEEE Access, vol. 9, pp. 10473 - 10497, 2021.
- [SCC-6] L. Dan, Y. Zheng , D. Wenxiu y M. Atiquzzaman, «A survey on secure data analytics in edge computing,» IEEE Internet of Things Journal, vol. 6, pp. 4946-4967, 2019.
- [SCC-7] M. Nikravan y M. H. Kashani, «A review on trust management in fog/edge computing: Techniques, trends, and challenges,» A review on trust management in fog/edge computing: Techniques, trends, and challenges, vol. 2022, 2022.
- [SCC-8] G. Zhipeng , W. Zhao, C. Xia, X. Kaile , Z. Mo, Q. Wang y Y. Yang, «A Credible and Lightweight Multidimensional Trust Evaluation Mechanism for Service-Oriented IoT Edge Computing Environment,» de 2019 IEEE International Congress on Internet of Things (ICIOT), 2019.
- [SCC-9] W. Tian , L. Hao , Z. Xi y X. Mande, «Crowdsourcing mechanism for trust evaluation in CPCS based on intelligent mobile edge computing,» ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, pp. 1-19, 2019.
- [SCC-10] K. N. Qureshi, A. Iftikhar, S. N. Bhatti, F. Piccialli, F. Giampaolo y J. Gwanggil, «Trust management and evaluation for edge intelligence in the Internet of Things,» Engineering Applications of Artificial Intelligence, vol. 94, 2020.

From DevOps to DevSecOps to DevPrivSecOps

- [DPSO-1] The Incredible True Story of How DevOps Gots Its Name. 2014. <https://newrelic.com/blog/nerd-life/devops-name>
- [DPSO-2] C. Ebert, G. Gallardo, J. Hernantes, and N. Serrano. 2016. DevOps. IEEE Software 33, 3 (2016), 94–100. Code: A54
- [DPSO-3] L. Leite, C. Rocha, F. Kon, D. Milojicic, P. Meirelles, “A survey of DevOps concepts and challenges”, ACM Computing Surveys, Vol. 52, No. 6, Article 127. Nov. 2019
- [DPSO-4] The Eight Phases of a DevOps Pipeline; <https://medium.com/taptuit/the-eight-phases-of-a-devops-pipeline-fda53ec9bba>
- [DPSO-5] Rakesh Kumar, Rinkaj Goyal, July 2020 “Modeling continuous security: A conceptual model for automated DevSecOps using open-source software over cloud ADOC”
- [DPSO-6] What are the Phases of DevSecOps?; Veritis; <https://www.veritis.com/blog/what-are-the-phases-of-devsecops/#02>

Distributed multiplane analytics

- [DMA-1] A. Clemm, M. Chandramouli and S. Krishnamurthy, "DNA: An SDN framework for distributed network analytics," 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, pp. 9-17, doi: 10.1109/INM.2015.7140271

- [DMA-2] Mortier, Richard and Haddadi, Hamed and Servia, Sandra and Wang, Liang, " Distributed data analytics," arXiv, 2012, doi: 10.48550/ARXIV.2203.14088
- [DMA-3] N. Kefalakis, A. Roukounaki and J. Soldatos, "A Configurable Distributed Data Analytics Infrastructure for the Industrial Internet of things," 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2019, pp. 179-181, doi: 10.1109/DCOSS.2019.00048
- [DMA-4] M. Taneja, N. Jalodia and A. Davy, "Distributed Decomposed Data Analytics in Fog Enabled IoT Deployments," in IEEE Access, vol. 7, pp. 40969-40981, 2019, doi: 10.1109/ACCESS.2019.2907808..
- [DMA-5] Y. Chang, X. Huang, Z. Shao and Y. Yang, "An Efficient Distributed Deep Learning Framework for Fog-Based IoT Systems," 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9014056.
- [DMA-6] A. Forestiero and G. Papuzzo, "Distributed Algorithm for Big Data Analytics in Healthcare," 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 776-779, doi: 10.1109/WI.2018.00015
- [DMA-7] P. Triantafillou, "Towards Intelligent Distributed Data Systems for Scalable Efficient and Accurate Analytics," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2018, pp. 1192-1202, doi: 10.1109/ICDCS.2018.00119
- [DMA-8] A. Cuzzocrea, "Advanced, Privacy-Preserving and Approximate Big Data Management and Analytics in Distributed Environments: What is Now and What is Next," 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020, pp. 1564-1568, doi: 10.1109/COMPSAC48688.2020.00-32
- [DMA-9] M. H. Isah, T. Abughofa, S. Mahfuz, D. Ajerla, F. Zulkernine and S. Khan, "A Survey of Distributed Data Stream Processing Frameworks," in IEEE Access, vol. 7, pp. 154300-154316, 2019, doi: 10.1109/ACCESS.2019.2946884.
- [DMA-10] A. -A. -R. Nayeem et al., "A Visual Analytics Framework for Distributed Data Analysis Systems," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 229-240, doi: 10.1109/BigData52589.2021.9671768
- [DMA-11] C. P. Filho et al., "A Systematic Literature Review on Distributed Machine Learning in Edge Computing," Sensors, vol. 22, no. 7, p. 2665, Mar. 2022, doi: 10.3390/s22072665.

Industrial approach to edge-cloud continuum in Industry (I4.0 and I5.0)

- [IECC-1] Report. PAVING THE WAY FOR A DATA-DRIVEN INDUSTRY DIGITALISATION. FIWARE. 2018.
- [IECC-2] Report. Digitising European Industry Working Group 2 Digital Industrial Platforms Final version. European Commission. 2017.
- [IECC-3] Web page. Smart Industry <https://www.fiware.org/about-us/smart-industry/>. FIWARE.
- [IECC-4] S. De Panfilis, S. Gusmeroli, J. Rodríguez and J. Benedicto, "FIWARE for Industry: A Data-driven Reference Architecture" in Enterprise Interoperability, Hoboken, NJ, USA, pp. 171-178, 2018.
- [IECC-5] Report. BOOST 4.0 Reference Architecture Specification. 2019.
- [IECC-6] X. Xu, Y. Lu, B. Vogel-Heuser and L. Wang, "Industry 4.0 and Industry 5.0 - Inception, conception and perception" in Journal of Manufacturing Systems, vol. 61, pp. 530-535.
- [IECC-7] A. Marguglio, M. Pistone, A. Ude, J. Soldatos, M. Rocker and J. F. Ruiz, "Digital Factory Alliance- A Reference Architecture for digital Zero-Defect Manufacturing". [Online] <https://digitalfactoryalliance.eu/wp-content/uploads/2022/09/A-Reference-Architecture-for-digital-Zero-Defect-Manufacturing.pdf>
- [IECC-8] International Data Spaces Association (IDSA), "IDS-RAM 4.0" in IDS-RAM, vol.4. [Online] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/main/documentation

- [IECC-9] Open Industry 4.0 Alliance, “Technical Solution Design Principles”. [Online] <https://openindustry4.com/download-center/>
- [IECC-10] M. Moghaddam, M. Cadavid, R. Kenley and A. Deshmukh, “Reference architectures for smart manufacturing: A critical review” in Journal of Manufacturing Systems, vol. 49, pp. 215-225.
- [IECC-11] Industrial Internet Consortium, “Industrial Internet Reference Architecture (IIRA)” in Industrial Internet Consortium, Needham, MA, USA.
- [IECC-12] Z.V.E.I., “The Reference Architectural Model Industrie 4.0 (RAMI 4.0)”, vol. 14. [Online] <https://www.zvei.org/en/press-media/publications/the-reference-architectural-model-industrie-40-rami-40>

Current existing standards related to aerOS

- [CES-1] Object Management Group, “Data Distribution Service (DDS), Version 1.4”. March 2015.
- [CES-2] Real Time Innovations, “Data Distribution Service (DDS) for Complex Systems,” RTI. [Online]. Available: <https://www.rti.com/products/dds-standard>. [Accessed: 25-Oct-2022].
- [CES-3] John S Rinaldi, “OPC UA - Unified Architecture: The Everyman's Guide to the Most Important Information Technology in Industrial Automation”, April 2016
- [CES-4] “OPC Unified architecture,” OPC Foundation, 26-Sep-2019. [Online]. Available: <https://opcfoundation.org/about/opc-technologies/opc-ua/>. [Accessed: 26-Oct-2022].
- [CES-5] M. Bjorklund, Ed, “RFC 7950 - The YANG 1.1 Data Modeling Language,” Aug-2016 [Online]. Available: <https://datatracker.ietf.org/doc/html/rfc7950>. [Accessed: 25-Oct-2022].
- [CES-6] ETSI, “ETSI GS CIM 006 - V1.1.1 - Context Information Management (CIM); Information Model (MOD0),” July 2019
- [CES-7] ETSI, “ETSI GS CIM 009 - V1.6.1 - Context Information Management (CIM); NGSI-LD API,” August 2022.
- [CES-8] Time-Sensitive Networking (TSN) Task Group, <http://www.ieee802.org/1/tsn>
- [CES-9] IEEE Standard 802.1AS-2020, “IEEE Standard for Local and Metropolitan Area Networks – Timing and Synchronization for Time-Sensitive Applications”, published June 2020.
- [CES-10] IEEE Standard 1588–2019, “IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems”, published November 2019.
- [CES-11] IEEE Standard 802.1Q-2018, “IEEE Standard for Local and metropolitan area networks: Bridges and Bridged Networks”, published July 2018.
- [CES-12] IEEE Standard 802.3-2018, “IEEE Standard for Ethernet”, published August 2018.
- [CES-13] IEEE 802.1CB-2017, “IEEE Standard for Local and metropolitan area networks-Frame Replication and Elimination for Reliability”, published October 2017
- [CES-14] IETF, Deterministic networking (DetNet). [Online]. Available: <https://datatracker.ietf.org/wg/detnet/charter/>. [Accessed: 28-Oct-2022].

Global Analysis of the European Research on the Edge-Cloud Computing Continuum

- [ORP-1] The Network Compute Fabric, Ericsson Technology Review, ISSN 0014-0171 284 23- 3360, June 2021.

- [ORP-2] I. Petri et al., "Autonomics at the Edge: Resource Orchestration for Edge Native Applications," in *IEEE Internet Computing*, v. 25, n. 4, pp. 21-29, 2021
- [ORP-3] A. Manzalini and N. Crespi, "An edge operating system enabling anything-as -a-service," *IEEE Communications Magazine*, 54(3), 62-67, March 2016
- [ORP-4] R. Debab et al. "Boosting the Cloud Meta-Operating System with Heterogeneous Kernels. A Novel Approach Based on Containers and Microservices." *Journal of Engineering Science and Technology Review* 11, 103-108, 2018.
- [ORP-5] T. Taleb, et al., "On multi-domain network slicing orchestration architecture & federated resource control," *IEEE Network*, 33(5):242 -252, Sep. 2019.
- [ORP-6] ETSI GR NFV-IFA 028, "Report on architecture options to support multiple administrative domains," Jan. 2018.
- [ORP-7] ETSI GS NFV-IFA 040, "Requirements for service interfaces and object model for OS container management and orchestration specification," Nov. 2020.
- [ORP-8] M. Bagaa, et al., "Collaborative cross system AI: Toward 5g system and beyond," in *IEEE Network*, 35(4):286 – 294, Jul.2021.
- [ORP-9] D. Sarddar, et al, Refinement of Resource Management in Fog Computing Aspect of QoS, *Int. Journal of grid and Distributed*, v. 11. no 5, pp.29–44, 2018.
- [ORP-10] A. Athreya, et al, "Designing for Self-Configuration and Self-Adaptation in the Internet of Things", in *Proc. 9th ICCCN*, Oct. 2013, pp. 585-592
- [ORP-11] D. Roca, et al, Fog function virtualization: A flexible solution for IoT applications, in: *2017 2nd Int. Conf. on Fog and MEC*, Valencia, 2017, pp. 74–80
- [ORP-12] W. Wong, et al., Container deployment strategy for edge networking, in: *Proc. of the 4th Workshop on, MECC '19*, Davis, California, 2019, pp. 1–6.
- [ORP-13] B. Donassolo, et al., "Fog Based Framework for IoT Service Provisioning," *2019 16th IEEE CCNC*, 2019, pp. 1-6
- [ORP-14] Z. Nezami, et al., Decentralized Edge-to-Cloud Load Balancing: Service Placement for the Internet of Things, *IEEE Access*, vol. 9, pp. 64983-65000, 2021.
- [ORP-15] M. Alam, et al., Orchestration of microservices for IoT using docker and edge computing, *IEEE Commun. Mag.* 56 (9) (2018) 118–123.
- [ORP-16] T. P. Raptis, et al., "Data Management in Industry 4.0: State of the Art and Open Challenges," in *IEEE Access*, vol. 7, pp. 97052-97093, 2019
- [ORP-17] M. Allen, D. Cervo, "Metadata Management", in book *Multi-Domain Master Data Management*, MK Publishers, 2015, 161-178.
- [ORP-18] The European AI on Demand Platform: <https://www.ai4europe.eu/>
- [ORP-19] I. Kholod, et al. "Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis", *Sensors*, 21(1):167, 2021
- [ORP-20] J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu, "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues," in *IEEE Access*, vol. 6, pp. 18209-18237, 2018.
- [ORP-21] Y. Yu, R. Chen, H. Li, Y. Li and A. Tian, "Toward Data Security in Edge Intelligent IIoT," in *IEEE Network*, vol. 33, no. 5, pp. 20-26, Sept.-Oct. 2019
- [ORP-22] W. Shi and S. Dustdar, "The Promise of Edge Computing," in *Computer*, vol. 49, no. 5, pp. 78-81, May 2016
- [ORP-23] Y. Ruizhe, et al. "Integrated blockchain and edge computing systems: A survey, some research issues and challenges." *IEEE Communications Surveys & Tutorials* 21.2 (2019): 1508-1532.

[ORP-24] https://xeniro.io/wp-content/uploads/2019/07/XENIRO_Lite_WP_latest.pdf

Edge-cloud technologies in robotics and manufacturing sector

[EMS-1] L. Da Xu, E. L. Xu, and L. Li, “Industry 4 . 0 : state of the art and future trends,” Int. J. Prod. Res. ISSN, vol. 56, no. 8, pp. 2941– 2962, 2018.

[EMS-2] M. Rüßmann et al., “Industry 4.0:Future of Productivity and Growth in Manufacturing,” Bost. Consult. Gr., no. **April**, p. 20, 2015.

[EMS-3] E. Negri, L. Fumagalli, and M. Macchi, \A Review of the Roles of Digital Twin in CPS-based Production Systems," Procedia Manufacturing, 2017.

[EMS-4] Cite: National Science Foundation (NSF), \Cyber-Physical Systems (CPS)," 2014.

[EMS-5] IEC, \IEC 62264-2: Enterprise-control system integration { Part 2: Objects and attributes for enterprise-control system integration," 2015

[EMS-6] International Society of Automation, Enterprise-Control System Integration Part 2 : Object Model Attributes, 2001

[EMS-7] P. Bernus, L. Nemes, and G. Schmidt, \Handbook on Enterprise Architecture," Strategy, 2003

[EMS-8] <https://doi.org/10.1016/j.promfg.2020.02.055> (it's mine, Imao)

[EMS-9] Gartner, Manufacturing Execution System, 2015

[EMS-10] A. Rajhans, S.-W. Cheng, B. Schmerl, D. Garlan, B. H. Krogh, C. Agbi, and A. Bhave, \An Architectural Approach to the Design and Analysis of Cyber-Physical Systems," in 3rd International Workshop on Multi-Paradigm Modeling (MPM), 2009

[EMS-11] F. Menge, \Enterprise Service Bus," 2007

[EMS-12] R. Perrey and M. Lycett, \Service-oriented architecture," in Proceedings - 2003 Symposium on Applications and the Internet Workshops, SAINT 2003, 2003

[EMS-13] D. Chappell, Enterprise Service Bus: Theory in Practice, 2004.

[EMS-14] D. Arne, \Enterprise Service Bus," JavaSPEKTRUM, 2005.

[EMS-15] M. Marian, \iPaaS: Diferent Ways of Thinking," Procedia Economics and Finance, 2012.

Edge-cloud technologies in maritime port sector

[EMP-1] P. Beaumont, “Cybersecurity Risks and Automated Maritime Container Terminals in the Age of 4IR,” in Handbook of Research on Information and Cyber Security in the Fourth Industrial Revolution, IGI Global, 2018, pp. 497–516.

[EMP-2] K. Rintanen and A. Thomas, “Container terminal automation,” 2016.

[EMP-3] I. de la Peña Zarzuelo, M. J. Freire Soeane, and B. López Bermúdez, “Industry 4.0 in the port and maritime industry: A literature review,” J. Ind. Inf. Integr., vol. 20, p. 100173, 2020.

[EMP-4] A. M. Martín-Soberón, A. Monfort, R. Sapiña, N. Monterde, and D. Calduch, “Automation in Port Container Terminals,” Procedia - Soc. Behav. Sci., vol. 160, no. Cit, pp. 195–204, 2014.

[EMP-5] N. Zrnić, Z. Petković, and S. Bošnjak, “Automation of ship-to-shore container cranes: A review of state-of-the-art,” FME Trans., vol. 33, no. 3, pp. 111–121, 2005.

[EMP-6] M. Etienne, S. Khan, and M. Eakambaram, “Modern ships and ports,” 2020.

[EMP-7] R. Kompany, “5G and MEC can significantly improve smart port operations,” 2019, Online: <https://www.analysysmason.com/Research/Content/Comments/5g-mec-ports-rma18/>

- [EMP-8] T. Zonta, C.A. Da Costa, R. da Rosa Righi, M.J. de Lima, E.S. da Trindade and G.P. Li, 2020. Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150, p.106889.
- [EMP-9] K. Koo, E. Cha, 2012. A Novel container ISO-code recognition method using texture clustering with a spatial structure window. *International Journal of Advanced Science and Technology*. 41
- [EMP-10] Y. Liu, Y. Want, J. Lv, M. Zhang, 2012. Automatic spreader-container alignment system using infrared structured lights. *Applied Optics*, 51(16)
- [EMP-11] A. Andziulis, T. Eglynas, M. Bogdevicius, T. Lenkauskas, M. Juis, 2016. Development of an adaptive intermodal container handling control subsystem based on automatic recognition algorithms. *European International Journal of Science and Technology*, 5(3).
- [EMP-12] iTerminals 4.0, “Application of Industry 4.0 Technologies towards Digital Port Container Terminals,” 2018., Online: <https://iterminalsproject.eu/>
- [EMP-13] Horizon 2020, “Capacity with a pOsitive enviRonmEntal and societAL footprInt: portS in the future era,” 2018., Online: <https://cordis.europa.eu/project/id/768994>
- [EMP-14] Horizon 2020, “Cyber preparedness actions for a holistic approach and awareness raising in the MARitime logistics supply chain,” 2019, Online: <https://cordis.europa.eu/project/id/833389>
- [EMP-15] MarineInsight, “Rolls-Royce Offers Ship Navigators A Bird’s-Eye View With Intelligent Awareness Game-Changer,” 2018m Online: <https://www.marineinsight.com/shipping-news/rolls-royce-offers-ship-navigators-a-birds-eye-view-with-intelligent-awareness-game-changer/>
- [EMP-16] R. Cardone, “The 5G Port of the Future,” 2020, Online: <https://www.ericsson.com/en/blog/2020/7/the-5g-port-of-the-future>
- [EMP-17] Huawei, “Working with China Telecom Huawei will Build the MEC Network of Ningbo Zhenhai Smart Refinery,” 2017.

Edge-cloud technologies in mobile machinery sector

- [EMM-1] “The NIST definition of cloud computing.” [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>. [Accessed: 02-Dec-2022].

Edge-cloud technologies in telecom operators sector (a usability perspective)

- [ETS-1] GSMA Intelligence, 9/2022, “5G for the enterprise: headway, hurdles and the horizon for operators”, <https://data.gsmainelligence.com/research/research/research-2022/5g-for-the-enterprise-headway-hurdles-and-the-horizon-for-operators>
- [ETS-2] Harmonizing standards for edge computing A synergized architecture leveraging ETSI ISG MEC and 3GPP specifications, July 2020, https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_wp36_Harmonizing-standards-for-edge-computing.pdf
- [ETS-3] Operators increase the efficiency of their networks to reduce costs , <https://on5g.es/en/las-operadoras-aumentan-la-eficiencia-de-sus-redes-para-reducir-costes/>
- [ETS-4] <https://www.openstack.org/use-cases/telecoms-and-nfv/>
- [ETS-5] <https://telco.vmware.com/>
- [ETS-6] Migration from Physical to Virtual Network Functions: Best Practices and Lessons Learned, <https://www.gsma.com/futurenetworks/5g/migration-from-physical-to-virtual-network-functions-best-practices-and-lessons-learned/>

- [ETS-7] GSMA 2022 Mobile Economy Report Europe (10/2022), <https://www.gsma.com/mobileeconomy/europe/>
- [ETS-8] GSMA report demonstrates policy action is needed for EU to achieve Digital Decade goals, <https://www.gsma.com/gsmaeurope/news/mobile-economy-europe-2022/>
- [ETS-9] Harmonizing standards for edge computing A synergized architecture leveraging ETSI ISG MEC and 3GPP specifications, July 2020, https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_wp36_Harmonizing-standards-for-edge-computing.pdf
- [ETS-10] <https://www.3gpp.org/>
- [ETS-11] Applications Enablement Standards in 3GPP, <https://www.3gpp.org/news-events/3gpp-news/sa6-app-enable>
- [ETS-12] System architecture for the 5G System (5GS) , 3GPP TS23.501, https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/16.06.00_60/ts_123501v160600p.pdf
- [ETS-13] Generic Network Slice Template, <https://www.gsma.com/newsroom/wp-content/uploads//NG.116-v7.0.pdf>
- [ETS-14] 5G Network Programmability (NEF), 3GPP TS 29.522 version 15.3.0 Release 15, https://www.etsi.org/deliver/etsi_ts/129500_129599/129522/15.03.00_60/ts_129522v150300p.pdf
- [ETS-15] Network Data Analytics Services (NWDAF); 3GPP TS 29.520 version 15.3.0 Release 15, https://www.etsi.org/deliver/etsi_ts/129500_129599/129520/15.03.00_60/ts_129520v150300p.pdf
- [ETS-16] Telco Edge Computing - Mapping requirements to standards, <https://www.3gpp.org/news-events/3gpp-news/mec-gsma>
- [ETS-17] Common API Framework for 3GPP Northbound APIs (CAPIF), 3GPP TS 23.222 version 15.3.0 Release 15, https://www.etsi.org/deliver/etsi_ts/123200_123299/123222/15.03.00_60/ts_123222v150300p.pdf
- [ETS-18] Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows 3GPP TS 23.434 version 16.4.0 Release 16, https://www.etsi.org/deliver/etsi_ts/123400_123499/123434/16.04.00_60/ts_123434v160400p.pdf
- [ETS-19] Architecture for enabling Edge Applications (EDGEAPP), 3GPP TS 23.558 version 17.3.0 Release 17, https://www.etsi.org/deliver/etsi_ts/123500_123599/123558/17.03.00_60/ts_123558v170300p.pdf
- [ETS-20] 3GPP TS 28.538 version 17.0.0 Release 17, Management and orchestration; Edge Computing Management , https://www.etsi.org/deliver/etsi_ts/128500_128599/128538/17.00.00_60/ts_128538v170000p.pdf
- [ETS-21] 3GPP TS 28.552 version 16.6.0 Release 16, Management and orchestration; 5G performance measurements http://www.etsi.org/deliver/etsi_ts/128500_128599/128552/16.06.00_60/ts_128552v160600p.pdf
- [ETS-22] 3GPP TS 28.554 version 16.7.0 Release 16, Management and orchestration; 5G end to end Key Performance Indicators (KPI), http://www.etsi.org/deliver/etsi_ts/128500_128599/128554/16.07.00_60/ts_128554v160700p.pdf
- [ETS-23] ETSI MEC, Multi-access Edge Computing (MEC), <https://www.etsi.org/technologies/multi-access-edge-computing>
- [ETS-24] ETSI MEC website, <https://www.etsi.org/technologies/multi-access-edge-computing>
- [ETS-25] ETSI GS MEC 009 V3.1.1 (2021-06), “Multi-access Edge Computing (MEC); General principles, patterns and common aspects of MEC Service APIs”, https://www.etsi.org/deliver/etsi_gs/MEC/001_099/009/03.01.01_60/gs_MEC009v030101p.pdf
- [ETS-26] OPG GSMA, <https://www.gsma.com/futurenetworks/5g-operator-platform/>

- [ETS-27] Operator Platform APIs,
https://www.3gpp.org/ftp/tsg_sa/WG6_MissionCritical/Informal_ConfCalls/ICC_20220110_eEDGEAPP/inbox/OPAG_Webinar_02Dec_v0_5.pdf
- [ETS-28] TEC GSMA, <https://www.gsma.com/futurenetworks/resources/telco-edge-cloud-october-2020-download/>
- [ETS-29] <https://www.gsma.com/futurenetworks/wp-content/uploads/2022/03/TEC-Forum-04-Open-Session-Slides-100322-v2.pdf>
- [ETS-30] tmforum, <https://www.tmforum.org/about-tm-forum/>
- [ETS-31] <https://www.mobileworldlive.com/featured-content/home-banner/gsma-chair-opens-door-to-tm-forum-collaboration/>
- [ETS-32] Becoming EDGY, https://www.tmforum.org/collaboration/catalyst-program/becoming-edgy/?_gl=1*4s5ri3*_ga*ODYyNzg4MDUzLjE2NjY3NjcwNzc.*_ga_W21R8NVK4E*MTY2Njg1NDQ5My4zLjEuMTY2Njg1ODI1Mi4wLjAuMA..&_ga=2.242381441.2073117244.1666767077-862788053.1666767077
- [ETS-33] tmforum, Open Digital Architecture, <https://www.tmforum.org/oda/>
- [ETS-34] GPP TS 28.541 version 16.6.0 Release 16, Management and orchestration; 5G Network Resource Model (NRM),
https://www.etsi.org/deliver/etsi_ts/128500_128599/128541/16.06.00_60/ts_128541v160600p.pdf
- [ETS-35] ETSI GR MEC 035 V3.1.1 (2021-06), “Multi-access Edge Computing (MEC); Study on Inter-MEC systems and MEC-Cloud systems coordination”,
https://www.etsi.org/deliver/etsi_gr/MEC/001_099/035/03.01.01_60/gr_mec035v030101p.pdf
- [ETS-36] ETSI White Paper No. 24, “MEC Deployments in 4G and Evolution Towards 5G”, February 2018,
https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp24_MEC_deployment_in_4G_5G_FINAL.pdf
- [ETS-37] ETSI White Paper No. 28, “MEC in 5G network”, June 2018,
https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf
- [ETS-38] ETSI MEC feedback on SDO mapping,
https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_WP_49_MEC-Federation-Deployment-considerations.pdf
- [ETS-39] tmforum, <https://www.tmforum.org/oda/about-open-apis/>
- [ETS-40] tmforum NaaS openAPI, https://www.tmforum.org/resources/specification/tmf909-api-suite-specification-for-naas-v3-0/?_gl=1*sb6rcz*_ga*ODYyNzg4MDUzLjE2NjY3NjcwNzc.*_ga_W21R8NVK4E*MTY2Njg1NDQ5My4zLjEuMTY2Njg1OTYxMC4wLjAuMA..&_ga=2.210490774.2073117244.1666767077-862788053.1666767077
- [ETS-41] Tmforum Sustainability topic, <https://inform.tmforum.org/topics/sustainability/>

Edge-cloud technologies in containerised data centres close to renewable energy sources

- [ERE-1] Albert Greenberg, James Hamilton, David A. Maltz, Parveen Patel, Microfoft Research, 2009. The Cost of a Cloud: Research Problems in Data Center Networks
- [ERE-2] FRANCO-GERMAN POSITION PAPER ON "SPEEDING UP INDUSTRIAL AI AND TRUSTWORTHINESS", 2021
- [ERE-3] David Mytton, Research Affiliate, Uptime Institute, 2021. Renewable energy for data centers.
- [ERE-4] <https://www.se.com/ww/en/work/solutions/for-business/data-centers-and-networks/>

[ERE-5] <https://www.dell.com/pl-pl/dt/solutions/modular-data-centers.htm#scroll=off&tab0=0>

[ERE-6] <https://www.vertiv.com/en-asia/solutions/>

[ERE-7] <https://www.kstar.com/ContainerDataCenters/index.jhtml>

[ERE-8]

https://www.cisco.com/c/dam/en_us/solutions/industries/docs/gov/Containerized_Data_Centers_Solution_Overview.pdf

[ERE-9] <https://www.gstatic.com/gumdrop/sustainability/247-carbon-free-energy.pdf>

[ERE-10] <https://www.starlingx.io/>

[ERE-11] <https://www.airshipit.org/>

A. Interviews

Topics	Key contributions
Interest towards aerOS	<ul style="list-style-type: none"> • The direct need, observed in the research field, for the Meta Operating System, that could actually cover the continuum of resources in the IoT cloud sector. The experience gained through research and transfer technology projects allowed to realise that the current reality shows a too extended and strict commitment to a certain and regular specific vendor, hardware, network and so on. Instead, aerOS proposes a much opener approach that takes into account interoperability and open-source trends, allowing these services to be deployed in different instances throughout the continuum. • The IoT Edge-to-Cloud continuum solution proposed in aerOS offers a great deal of optimisation of resources and services (e.g., for the mobile machinery system, it makes in-vehicle edge nodes, smart sensors and network components more efficient). • Compared to other projects on the IoT ecosystem, aerOS is a natural continuation of the research, as it goes a step further on the evolution of the domain. In fact, aerOS provides an accurate forecast for the future developments of the Industrial Edge, being already in its own way a Meta Operating System and allowing to experience the combination with advanced networking technologies (e.g., 5G or TSN). • From the perspective of sustainability for customers like factories, the solution will allow to make more targeted changes to the operational structure by adapting to the existing situation without having to destroy everything and rebuild from scratch. • Through an internal big data management perspective, investigate and support the proper movements of some components of the solution from Cloud to Edge. • The interest in the cloud continuum domain of aerOS, considered the importance that virtualisation, softwarisation and cloud computing in general play in 5G and mobile communications technologies: aerOS will help integrating the current 5G systems and providing services to data fusion. • The relevance of aerOS technology in applications for data protection processing in Industry 4.0 where low latencies are requested. The aerOS project is on the right path to provide the ability to operate on different platforms and different levels of computing and to make the use and computing of the information more transparent and less cumbersome. • The need to enable the deployment of intelligence in appliances such as mobile machines and to optimise their work amid limited resources.
aerOS potential impact in enterprises <i>(in particular, the impact of an edge cloud solution)</i>	<ul style="list-style-type: none"> • Potentially, a technology like aerOS could bring to tech transfer contracts with the private sector and to the creation not only of new research groups, but also to spin-offs and start-ups. • For stakeholders with industrial customers (e.g., construction and agriculture), a solution developed through real-world use cases like aerOS

	<p>will help to expand their market share by keeping the solutions provided relevant and updated thanks to the optimisation of resources and services. In fact, Cloud services and/or fog and edge computing will enable more efficiency in big data analytics and remote data management, and will help deploying advanced orchestration methods to be combined with their current connectivity and cloud solutions.</p> <ul style="list-style-type: none"> • As an Industrial Edge kind of solution, aerOS will bring benefit to parallel Industrial 5G products and to crane technologies; furthermore, to the manufacturing industry domain and to the smart infrastructure domain. • The data movement from the cloud to the edge will allow to predict potential accidents in the terminals (and in general, predictive maintenance) and to acquire better protection from economic crimes, having not to wait for the management delays of the cloud and permitting as a response fast and decisive actions. • As a newly introduced service innovation of cloud continuum, aerOS will be a trigger for Europe to optimise either the performance of existing services and products and to introduce new ones. • aerOS will provide an impact both from a technological point of view for companies related to infrastructure virtualisation and network continuum that might adopt its outcomes (e.g., the continuum), and from an operational and business management point of view for companies which have a great deal of data produced that need to be processed locally but orchestrated from a cloud perspective, through the notion of operating system as a continuum across the cloud. • For the industrial phenomenon of “servitisation”, that consists in selling services around a physical asset, aerOS platform will provide value added services like smart control, active monitoring, energy efficiency, remote configuration and continuous Technical Support assistance, thanks to its use of the Edge. • In Academia, an Edge-to-Cloud solution as aerOS would support the need to expand the horizons of research, improving their capacity in doing so (not only from an economic perspective) with more targeted tools.
--	---

Market trends	<ul style="list-style-type: none"> • Compatibility with already existing cloud operating systems and relative improvement. • Positive environmental impact of proposing solutions that allow to reduce carbon footprint in the IoT domain. • Data security, data reliability and protection-oriented solutions. • Presenting a continuum, a common layer in services and topologies that are quite disparate in logic and topology; expanding the notion of the computing resources from a locality to a more distributed network connected infrastructure. • Solution-oriented protocols, work flows and attractions. • Resilience and flexibility in implementing faster responses to industrial requirements and unplanned events. • Increasing regulatory requirements for cradle-to-grave monitoring and recording of work functions (for example in agriculture).
Competitors/Comparable	<p style="text-align: center;">Competitor (to edge cloud – IoT – Data)</p>

	<ul style="list-style-type: none"> • GAIA-X: a project promoted by the EU for the development of an open source-based European data infrastructure. When operational, GAIA-X will enhance Europe's digital sovereignty maintaining the privacy of European data while allowing users to work simultaneously with multiple clouds. • ZeroNet: a Hungarian decentralised web-like network of peer-to-peer users. It is built in Python and is fully open-source. It uses bitcoin cryptography and trackers from the BitTorrent network to negotiate connections between peers. • Huawei Open Source, evolved from Huawei CarbonData program and sponsored by the Apache Software Foundation. • NEC Electronics GmbH open-source implementation solution included in the work of firmware. • Cloud services provided by Balena.io (Greece, United Kingdom).
	<p style="text-align: center;">Comparable (to edge cloud – IoT – Data)</p> <ul style="list-style-type: none"> • Arrowhead open-source service-oriented architecture and tools that allow the orchestration of microservices in a safe way, by LTU (Luleå University of Technology, Sweden). • ROS - robot operating system, specifically focused on robot management. Despite its name, it is not an operating system (OS) but a set of software frameworks for robot software development and provides services designed for a heterogeneous computer cluster such as hardware abstraction, low-level device control, implementation of commonly used functionality, message-passing between processes and package management.
Alternative solutions for data management/ processing/ analysis	<ul style="list-style-type: none"> • New communication protocols, such as: OPC Unified Architecture and Zenoh (Zero Overhead Network Protocol). • Commercial cloud-based services and platforms like Google Cloud or Microsoft Azure. • Open-source data management solutions (NoSQL), such as: Elasticsearch and MongoDB. • Industrial IoT solutions, such as: MindSphere by Siemens. • Industrial Engineering solutions by Dassault Systèmes.
Entry barriers	<ul style="list-style-type: none"> • Ethical and legal issue, due to the uncertainty of future regulations and future mandates makes stakeholders and entities reluctant to invest, a fact that prevents technologies from advancing at the rhythm they should. • The reluctance to share information and data, even knowing that they are supposed to stay on-premise. So, privacy and security concerns related to open platforms. • Slowdowns and obstacles in communication with different departments of large entities, with too long procedures that discourage investors. • The technological lock-in of digital solutions, where processes in an enterprise are locked to a specific solution, making the transition to a new one more difficult, even if it were more effective. It is combined with technological inertia and reluctance to take risks. • User understanding and acceptance, combined with the learning curve of a new and disruptive solution like aerOS; the sustainability of AI-based solutions.

	<ul style="list-style-type: none"> • Business strategies in many sectors that do not include open solutions. • The concern about the potential lack of connection to and implementation of reference models and existing architectures used by significant industrial players on the market.
Barriers in the adoption of an EU funded project solution	<ul style="list-style-type: none"> • The generally encountered market reluctance against products or services delivered from European projects, based on concerns about the feasibility, functionality and applicability of the solutions beyond the limited and controlled pilots where they were tested and developed. The use case scenarios are perceived as non-completely realistic because of their constraint and limited environments: for example, the security issue, which is pivotal in real life, is perceived by potential customers as too simplified in the laboratory and therefore not properly dealt with. • Since the prototypes from funded R&D projects are usually matured for sale after the project within 3-7 years depending on the application domain, the adoption issue does not concern encountering internal company reluctance, but further investments and time needed to advance the prototypes to a product level, and rapidly changing customer requirements. The real product finalisation goes beyond the European project scope: the funded project normally allows to go up to TRL 4-5 for RIA and a bit higher for IA. In order to be put on the market, a product needs normally another couple of design and manufacturing phases, sometimes certification steps, marketing campaign, dedicated customer documentation, and so on. • In certain fields such as ports, the market reluctance has not been clearly associated either with the “limited” European project development of the solution or with the general technological inertia in front of new and different digital solutions. These new solutions are often not acknowledged in their disruptive significance, and therefore not considered necessary and worth of such a big investment as the one they require to be integrated in the customer own system. • A partner encountered a real market reluctance against the European funded solution in the first couple of years after the project conclusion, when it was not considered properly tuned.
aerOS unique selling points	<ul style="list-style-type: none"> • aerOS will build a new Operating System for managing the cloud computing environment. Different clouds, from the Far Edge to the central Cloud, will have a single coordination centre. • The specific type of Edge Computing used in aerOS is a unique solution in Europe, since it is not currently being used anywhere. In fact, unlike competitor solutions, aerOS is a proper and complete open-source interoperable Edge continuum solution, that deals with the whole ecosystem and not only with specific challenges. • aerOS will expand the notion of the Operating System from a single hardware to a network and it will make it more user-friendly. • While creating a continuum between computing layers, aerOS will include several capabilities in the configuration and customisation of the computing environment, making it ready to operate on a wide range of platforms. The solution will provide not only implementation, but proper configuration, management and customisation of the orchestration. • aerOS solution will adapt to different use cases while still providing inter-compatibility to all the existing users, offering flexibility and scalability for different sectors. It will be deployed in heterogeneous scenarios, facing a vast number of different requirements and concerns of the concrete world of cloud edge, IoT and all the related fields, and demonstrating several functionalities in several verticals.

	<ul style="list-style-type: none"> • Different perspectives and needs will be incorporated in a Meta Operating System across the Continuum: for example, the visions of telecommunication operators, of properly Cloud-based SMEs, of Edge hardware providers, of Academia, and more. • Some of the aerOS outputs may exist on their own and therefore be incorporated into existing or in development products, improving their functionality.
Tech trends	<ul style="list-style-type: none"> • Bringing computation and intelligence to the Edge. • Edge Computing and hybrid cloud orchestration. • For Edge Cloud management and IoT segments, lightweight operating systems. • Digital Twin technologies. • Data Spaces technologies. • Modular and scalable in-vehicle platforms and modules. • Big Data, in the meaning of data-driven preventive and predictive maintenance. • Autonomy (for single vehicle or connected and cooperative swarms of vehicles).
How companies work with unstructured data and where data are stored	<ul style="list-style-type: none"> • For research institutions, since the amount of data is not too wide, the focus mainly shifts to infrastructure, architecture and algorithmic site; when it comes to data itself, the partners do researches on all the aspects related to semantics, to interoperability and to stream data processing. Concerning unstructured data, they have experience with natural language processing and different techniques used in it. They also deal with processing of images, that is to say preparing a model and also pre-processing images. • For partners in the industrial field, they have plenty of standardisation efforts (e.g., W3C Web of Things - WoT) in the optimisation scenario, in addition to MES systems (Manufacturing Execution Systems) to control from the SAP the data combined with customer queries going down to the factory. • For partners who are newcomers in Big Data management, the key to approach the matter is their expertise with the infrastructures necessary for handling the data, e.g., virtualisation environments like clouds. • For car manufacturer partners, their method is to share information from the engineering phase, in order to allow to process data and to understand how each component has been built. In terms of dealing with highly diverse and quite large volumes of data (a quite common phenomenon), usually the information is not maintained by a single system, and, in some cases, not even by the same entity, so to have mechanisms to effectively exchange or share that information is a significant issue they have to deal with. • As a solution for heterogeneous data management, a holistic and comprehensive platform that provides insights to terminal operators linking the work that the terminal operating system is managing with the telemetry of the crane or with the respective location.
How companies work with IoT ecosystems	<ul style="list-style-type: none"> • A partner presented an end-to-end IoT system as part of its current portfolio, offering an on-machine gateway, connectivity service and cloud back-end in a single turn-key solution. The solution currently supports more than 1.000 machines from around 20 customers.

	<ul style="list-style-type: none"> • A partner is interested in building an ecosystem where third party device manufacturers are engaging with in the perspective of multiple ecosystems intertwined. • A partner worked in an Inter IoT project concerning interoperability of IoT platforms on different levels. They were responsible for designing solutions for mapping different data, semantic translation and lifting of data (dealing with different syntaxes coming from different platforms). • The same partner is working in an Assist IoT project, being active in the area of data management (developing tools for interoperability, mapping, syntactic and semantic translation) and in the area of application of edge computing on different devices. They are using a software to philtre data (sharing the data and mapping the data model in order to make the data more understandable) and all the necessary equipment to gather information and measurements relating to workers safety. • In order to avoid terminal operations for decision making (incapable of producing quantitative data concerning the decisional background), a partner started another European project with the platform Inq-ITS, where they tried to integrate the data that the cranes were publishing, but not storing anywhere. As a device, they chose a specific IoT gateways from Siemens brand with a user-friendly interface and in which they were able to connect and link the data that were received from a specific data source. • A partner mentioned their previous experience in handling remote devices through different networks and IoT devices both in Smart City solutions, in Agriculture solutions and Smart Energy solutions, to remote control and remote management and remote recording, to gather the metrics and combine them for results-based decision-making. • A partner mentioned both their more traditional IoT devices, e.g., sensors for machine for maintenance operation of the equipment (IoT systems for predictive maintenance) and their less traditional IoT systems, e.g., a 3D scanner, a type of sensor developed by their group, that is working even in the design and electronics of the camera and all the different components inside of it and in pre-processing all this video information. The aim is to generate the point clouds and to manage all that information for the type of use case. They use an IoT hub where to collect all the information from the sensors, to philtre and to pre-process the data in the edge before sending to the cloud. The partner has used this type of systems as well for remote configuration and even operation of equipment.
Main benefits of edge cloud continuum system in EU	<ul style="list-style-type: none"> • Since Europe has already lost the race for cloud technologies against America or Asia, the race for creating open tools and standardising IoT ads and related deployments represents a too significant opportunity for European Countries to stay competitive if they reach the goal on time, also with solutions such as aerOS. The Edge to Cloud continuum system is pivotal for European non-dependence, sovereignty, and for a stronger position of European industry in the global market (including the whole value chain, e.g., technological components, systems, and so on). • It is really important for Europe to introduce the cloud continuum in order to optimise either the performance of existing services and products or to be the triggering point that will help the developers to introduce into the market innovative solutions, new services and new products. The cloud continuum will be even more revolutionary than the cloud computing was when introduced.

B. Focus groups

Topics	Key contributions
Interest towards aerOS	<p>SMART BUILDING</p> <ul style="list-style-type: none"> Partners underlined how the edge cloud momentum is an opportunity for the telecommunication operators to monetise their investments in 5G. It is crucial for them to observe the transformation that their clients are making and the vertical industries that they are mandating, in order to provide solutions and services that will properly respond to their needs. A digital transformation is essential for the partner themselves, as they need to reduce their Operational Expenditure (OpEx) and be energy efficient in their enterprise buildings located everywhere around the different countries. The development of more targeted services and data processes is also the main goal of the partners who wish to integrate their applications with the smart building system, which is spreading rapidly, making it more user friendly through a better visualisation of the data for the end users. The overall concept of aerOS with its cloud continuum (especially for the IoT) is something that will help the partners to redesign their existing IoT infrastructure, by using the LoRa WAN technology. All partners agreed on the fact that aerOS project will provide a unique solution coming from the joint minds of all the actors involved, with their different expertise and know-how and direct impact on the various use cases. <p>RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> Through aerOS, a better user experience can actually be achieved: the partners hope to use the solution in their containers, their micro data centres, and also to prepare better distributed processing on edge clouds. Energy creator companies desire to propose themselves on the market not only as providers of energy sources but also of the electronic energy loads necessary to connect the aforementioned sources to customers. Thus, energy could be produced by customers, managed together with them and be used locally, avoiding energy transfer to the network. <p>MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> aerOS will be an interesting solution to experiment the introduction of AI technologies in the partners facilities. Some partners wish to gain experience in the European project environment in order to boost their existing line of businesses of cluster-based supercomputing and to create new horizons as testing companies for validation purposes. <p>PORT CONTINUUM</p> <ul style="list-style-type: none"> The main interest from partners is to create a technical know-how in order to be prepared for the upcoming technological solutions that customers will soon require. In particular, the partners are interested in learning more about being at the forefront of Cloud Edge orchestration and distributed data processing systems. Furthermore, aerOS project deals in an innovative way with all the main technological topics of the future, AI technologies, communication systems and smart systems, IoT technologies and smart data processing. <p>MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p>

	<ul style="list-style-type: none"> The partners main interest consists in enabling the deployment of intelligence in mobile machines and in optimising their work amid limited resources. Furthermore, they wish to be able to examine the impact that aerOS and similar technologies would have on sustainability in farming and reducing energy consumption. The partners are interested in the fact that aerOS will allow in-vehicle edge nodes to interact with different smart devices, networking components, and computing continuum, which is currently challenging with limited resources.
aerOS potential impact in enterprises (in particular, the impact of an edge cloud solution)	<p style="text-align: center;">SMART BUILDING</p> <ul style="list-style-type: none"> A technology such as aerOS is leading the way to more innovative and fast changes in the Edge. In this perspective, and since the Edge deployments are very important for 5G implementations by the partners and for their approach towards the enterprise utilisation of the 5G capabilities (e.g., slicing private networks and so forth), the partners believe that aerOS will also support the business value proposition of their related applications and services provided to customers. Therefore, the enterprise, the business and the B2B business propositions will be inspired by the capabilities that will be provided by the aerOS private networks, since all these elements fit well into the Edge Cloud continuum of aerOS, that will help the internal digital transformation the partners are currently undergoing giving them a new selling point in being versatile towards new environments. aerOS is also regarded as efficient in the smart building operation and maintenance.
	<p style="text-align: center;">RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> The main impact will be the opportunity to find better solution to optimise distributed data processing, both in terms of timing and costs, also with a perspective of Green Edge processing. Compatibly with the legislative changes that will allow the solution to be adopted on a large scale, thanks to aerOS it will be possible to provide customers with new market horizons for their energy. In fact, it will not be necessary to deal with the costs coming from sending energy to the grid or with the fact that the current market is saturated. Without load limitations, energy could be sold it locally to the other companies, for example to data centres or electrolyzers.
	<p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> The introduction of AI technologies in the partners facilities will have a fundamental impact on all the company areas, especially the project one: all of them, together with the net of partners and customers, will take advantage from the knowledge coming from the European network environment, both for demonstrator applications, training and educational contents and data sharing with suppliers. Another crucial impact is represented by the fact that aerOS would boost the supercomputing line of business that the partners currently possess: the cluster supercomputing area will benefit from having a new technology implemented, that also may inspire the national governments to commission new and innovative technological projects to the partners.
	<p style="text-align: center;">PORT CONTINUUM</p>

	<ul style="list-style-type: none"> • Having a demo of an innovative solution such as aerOS will incentivise investors in investing more into the development of the final system, showing all the benefits that the partners will be already gaining. Furthermore, aerOS will improve the technical know-how of local realities, supporting the maintenance of machineries, together with security, technological and operational areas of the business. • For universities and research entities, the know-how acquired from the use case scenario will have a significant impact on the research about the maritime sector in general, improving both data collection, data storage and data analysis processes. <p style="text-align: center;">MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> • The fact that aerOS applications can be implemented in different industrial sectors like farming, construction, and forestry, brings valuable input through smart controls, and hence enhances the sustainability of farming. Current infrastructure of the partners, such as agricultural mobile machine technologies and related services, including precision farming, test fields, prototype construction machines, and wired as well as electrical mobile machines will benefit from the edge-cloud solutions that come with aerOS.
--	--

Market trends	<p style="text-align: center;">SMART BUILDING</p> <ul style="list-style-type: none"> • Cloud computing in the Smart Building Environment. Smart Cities market includes trends such as: cameras analysing traffic and controlling traffic lights (need for them to act in a coordinated matter), smart monitoring and remote control of public infrastructures. • Focus on energy efficiency and sustainability together with operational and OpEx reductions. <p style="text-align: center;">RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> • Green Edge Processing, meaning Edge Cloud Processing connected to renewable energy sources, focusing on sustainability and cost reduction. <p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> • Connection to the smart city market, intercommunication among smart building and smart city devices. • Connection to the agriculture market. • Introducing AI technologies in all the phases of the manufacturing process, also considering the supply chain management. <p style="text-align: center;">PORT CONTINUUM</p> <ul style="list-style-type: none"> • Digitalisation and automation in ports. • Providers supporting more than one single standard for connectivity, in order to improve the availability of equipment for developers.
----------------------	---

	<p>MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> Investigating the role of different Artificial Intelligence and Machine Learning techniques in different business and applications.
Market features	<p>SMART BUILDING</p> <ul style="list-style-type: none"> A partner reported that aerOS will respond to the market demand for an Edge processing which does not centralise to unknown Cloud operators, in addition to a IoT environment closer to end users. A partner reported that what is actually missing on the market or could be significantly improved thanks to aerOS is the presence of a higher layer in the architecture, meaning to be able to provide more software components in the service as a whole and in the intelligence management. <p>RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> A partner noticed the almost complete absence of comprehensive online Edge processing solutions like aerOS, that should be necessary for example for satellite images processing. Furthermore, edge components like micro data centres in the existing solutions are commonly connected to standard energy sources and are mainly used for disaster recovery. <p>MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> What is missing and requested by the manufacturing market is an interoperable system that allows customers not to change their data structures and data formats. It has been reported the need of having an orchestration not only of the data, but also of the components among each other, the ability to intercommunicate and interoperate between different machines and different types of information. The aspects of data security and data storage are already covered by existing solutions, but it necessary to implement solutions which are more elastic and flexible about data. <p>PORT CONTINUUM</p> <ul style="list-style-type: none"> A partner, which represent a small terminal that usually do not test very innovative technologies, reported to only have been using a custom software that their Mother Company developed and to not have tested any other platforms. Most of the technologies they require are available on the Market. A partner pointed out that the market requires to implement features such as: data security, data privacy and data resilience. <p>MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> The resources of current systems, such as connected mobile machinery, are being pushed to their limits, especially in tasks like data access and processing, ensuring data privacy and security or providing continuity to the Cloud.

Competitors/Comparable	<p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> • The ICOS (IoT to Cloud Operating System) European project. Like aerOS, the intended solution provides: device heterogeneity, continuum virtualisation, service orchestration, meta- OS, AI components. Like aerOS, it is tested on the following fields: Agriculture, Energy, Automotive, Transportation and Mobility (which aerOS will deal with thanks to the upcoming open calls). Unlike aerOS, it is not tested on: Logistics, Industry 4.0, Smart Cities and Health (the latter with the upcoming open calls of aerOS). • The project NEMO (Data processing and communication platform) will develop the first integrated sensing data platform for noise and exhaust emission measurements for individual vehicles. Like aerOS, the intended solution provides continuum virtualisation, service orchestration, meta-OS and AI components. Unlike aerOS, it does not provide device heterogeneity. Like aerOS, it is tested on the following fields: Agriculture, Energy, Automotive, Transportation and Mobility (which aerOS will deal with thanks to the upcoming open calls), Industry 4.0 and Smart Cities. Unlike aerOS, it is not tested on Logistics and Health (the latter with the upcoming open calls of aerOS). • The FLUIDOS (Flexible, scaLable secUre and decentralIseD Operating System) European project. Like aerOS, the intended solution provides service orchestration and meta-OS. It does not provide device heterogeneity, continuum virtualisation and AI components. Like aerOS, it is tested on the following fields: Agriculture, Energy and Logistics. Unlike aerOS, it is not tested on: Automotive, Transportation and Mobility (which aerOS will deal with thanks to the upcoming open calls), Industry 4.0, Smart Cities and Health (the latter with the upcoming open calls of aerOS). <p style="text-align: center;">PORT CONTINUUM</p> <ul style="list-style-type: none"> • Prodevelop. • TwinSIM project from Germany. <p style="text-align: center;">MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> • NaLamKI: a German research project, that works on developing AI services for use in agriculture, evaluating data from conventional and autonomous agricultural machinery, satellites and drones, and finally combining them in a software service platform and make the results accessible via open interfaces. • DEMETER: a European project with partners from 16 different Countries. One of its main objectives is transform the agricultural sector by building solutions on an array of digital technologies: Internet of Things, Earth Observation, Big Data, Artificial Intelligence, and of digital practices: cooperation, mobility, and open innovation.
-------------------------------	---

Entry barriers	<p style="text-align: center;">SMART BUILDING</p> <ul style="list-style-type: none"> All partners agreed on data privacy, data security and data handling: when data stay locally (it is supposed only in the building), it makes their availability impossible to not related people, making data controlling mechanism easier, in reverse to what happens in a more cloud-based environment, that is therefore regarded with suspicion by the market. <p style="text-align: center;">RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> The main barrier from the market point of view is the law. The partners are planning aerOS solution not only for their own countries, but to distribute it abroad. It has reported to be frequent in several countries the presence of legal restrictions against the connection to renewable energy sources, making possible only to connect micro data centres directly. <p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> All partners agreed on data security in this use case, too. They also mentioned the potential lack of interoperability and user friendliness as significant barriers. Therefore, the reluctance against digital solutions and technological inertia, together with lack of heterogeneity of data, that makes them difficult to manage. In fact, data management algorithm should be easy to be managed by the end user and be customised on the basis of the situation in which it has to be used. Lack of perception of the solution as really disruptive, not just an improvement of existing solutions. Lack of vision of the profitability of the technology that justifies the investment. <p style="text-align: center;">PORT CONTINUUM</p> <ul style="list-style-type: none"> All partners agreed on data security in this use case, too. Furthermore, for smaller port terminals, a barrier is represented by the lack of technical abilities of the team in dealing with new disruptive technologies. This kind of technical barrier is not perceived as present in bigger terminals. <p style="text-align: center;">MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> An open platform is commonly considered risky with regards to privacy and security of information related to customers
Barriers in the adoption of an EU funded project solution	<p style="text-align: center;">RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> Partners reported to be able to express only the general opinion that the solutions coming from a commercial environment since the beginning are considered better from a selling point of view to be commercialised. On the contrary, European funded solutions are developed mostly in laboratory. <p style="text-align: center;">MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</p> <ul style="list-style-type: none"> The partners reported to have experience in the sale of European funded solutions and not to have encountered any particular reluctance on the market.

aerOS unique selling points	<div data-bbox="826 192 1091 226" data-label="Section-Header"> <h3>SMART BUILDING</h3> </div> <div data-bbox="528 244 1444 412" data-label="List-Group"> <ul style="list-style-type: none"> • aerOS represents a new paradigm also compared to smart building vendors, because it allows the dynamic of placement of the employees in the smart building environment so that to provide energy efficient solutions. aerOS foresees an autonomous handling of smart buildings while at the same time maximising the resources and the profits. </div> <div data-bbox="722 461 1197 495" data-label="Section-Header"> <h3>RENEWABLE ENERGY SOURCES</h3> </div> <div data-bbox="528 512 1444 784" data-label="List-Group"> <ul style="list-style-type: none"> • aerOS is the only solution actually connected to renewable energy sources. Furthermore, its direct connection to the energy source will allow to avoid some costs for energy transferring, making the energy price lower. • aerOS is reported to be a quite unique solution from the perspective of batch processing and security, having a lot of smaller data centres as an edge component allowing to proceed with batch processing in a safer way even if any of the micro data centres should not be working. </div> <div data-bbox="679 833 1236 864" data-label="Section-Header"> <h3>MANUFACTURING AND PRODUCTION</h3> </div> <div data-bbox="528 884 1444 1458" data-label="List-Group"> <ul style="list-style-type: none"> • The main selling point of aerOS is the need of having an orchestration and the ability to intercommunicate and interoperate between different machines and different types of information. • The fact that aerOS comes as a European Standard is its biggest selling point. Having aerOS as a public, reference standard will shape how future European technology will be developed. Having a technology standardised by UE makes implementing such a technology a much safer and trustable decision. Adopting a standardised technology comes with a great set of benefits: knowing that other European companies use the same technology, which enables intercommunication and interoperability among companies or entities (e.g.: internet); long-term technological support and maintenance. technology evolution: if a technology is standardised and vastly used in society it is likely to evolve into improved iterations (e.g.: Ethernet, Wi-Fi), for everyone's benefit; a proprietary technology (contrary to a UE-standardised one) may end support shortly and may not ever evolve due to potential lack of users - this discourages companies. </div> <div data-bbox="815 1507 1101 1538" data-label="Section-Header"> <h3>PORT CONTINUUM</h3> </div> <div data-bbox="528 1559 1444 1796" data-label="List-Group"> <ul style="list-style-type: none"> • aerOS will facilitate and simplify the development of applications that go across Cloud, Edge and Far Edge, including clear API and, good standards automation in different aspects, both in terms of resource management and data movements. • Unlike competitors which are too generic and not specifically targeted, aerOS will be tested in several use case scenarios becoming a customised solution for several different sectors. </div> <div data-bbox="585 1843 1329 1912" data-label="Section-Header"> <h3>MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION</h3> </div> <div data-bbox="528 1930 1444 2000" data-label="List-Group"> <ul style="list-style-type: none"> • As organisations focusing on agricultural solutions, the partners identified aerOS main selling point in its care about sustainability of resources. </div>
------------------------------------	--

Tech trends	<p style="text-align: center;">SMART BUILDING</p> <ul style="list-style-type: none"> • Edge computing involved in the delivery of 5G technologies, in the data transfer and digital transformation of companies. <p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> • Cloud solutions vs on-premises solutions. • A common platform (both parties using the same technology) for manufacturers as suppliers and end customers, where to exchange data from IoT of the production to the end user side. <p style="text-align: center;">PORT CONTINUUM</p> <ul style="list-style-type: none"> • Communication medium support among the components: flexibility in machine-to-machine data communication with medium going from Wi-Fi to 4G/5G. • Standards for connectivity, implementation of protocols compatible with spywares.
How companies manage data	<p style="text-align: center;">SMART BUILDING</p> <ul style="list-style-type: none"> • Several partners (e.g., telco operators) reported to have been and to be frequently audited about their data management compliance with GDPR: they have to ensure that all the collected data do not span outside the European Union territory. So, they have to follow strict security protocols, avoid any partnership with high-risk suppliers and make sure than any data handling is actually following pseudonymisation and anonymisation techniques. <p style="text-align: center;">RENEWABLE ENERGY SOURCES</p> <ul style="list-style-type: none"> • A partner mentioned that, being data storage a crucial part of their activity, they are predominantly basing on shelf solution. They store all the downloads and the satellite images on daily basis and deliver this material to customers. Databases have significant dimensions. • A partner mentioned to possess a data analysis and data storage system in their company, connected to the maintenance service and to a surveillance centre. They send the data to customers, too. <p style="text-align: center;">MANUFACTURING AND PRODUCTION</p> <ul style="list-style-type: none"> • A partner reported that in their company, there is not a centralised data storage system, but every area has its own independent system. Nonetheless, they have a cloud system that store the data that are related to their aerOS pilot. • It also emerged that the presence of several different cloud tools to organise the data from their facilities or from demonstrators. For example, it has been mentioned Microsoft Azure for a demo platform and a new cloud solution from Schneider Electric where to summarise the data and display them. • For public institutions, it has been mentioned having CPD (Consumer Purchase Data) deployed, which gathers all the information and all the computation for the public administration, public health and also public

entities of the government. All their data storage happens in the CPD, which they have direct access to.

PORT CONTINUUM

- A partner reported that all the operational data, that means cargo data and anything related to the terminal operating system, are currently stored locally on their own servers. Technical data related to maintenance and maintenance procedures are stored on the Cloud, and the same happens with any administrative data as in HR data or health and safety data, training data. Security data are split halfway being on the cloud and on premises.
- A University partner has its own data centre where all the operational, financial and teaching data storage, computer and network equipment are kept.

MACHINERY FOR AGRICULTURE, FORESTRY AND CONSTRUCTION

- Customer related services are based on Amazon Web Services (AWS).

C. Written interviews

Topics	Key contributions
Use of external cloud infrastructure_Potential internal barriers	<ul style="list-style-type: none"> • A partner reported their normal use of One Drive by Microsoft, noticing the initial hurdle due to the compliance with internal IT policies for any cloud solution. • For Universities and Research Entities, the main potential hurdle could be represented by internal policies about opening networks to the WAN. • While noticing the absence of issues towards the acceptance of the deployment of operational applications and services in a cloud infrastructure, more than one partner recognised that there is the need to add some rules inside the firewall for the data coming from external sources, with all the relative costs. • A consideration that emerged has been the fact that migrating operational activities to the cloud would require training of personnel, digitalisation of records kept in physical form, and a particular care towards privacy of sensitive identifiable Information. • It has been pointed out that relinquishing the control to third party services and storing data and other applications in the cloud could provoke legal and regulatory concerns. Efficient ways would have to be determined in order to keep the organisation in step with compliance requirements, such as anonymisation tools and GDPR mechanics. • A significant hurdle for several less digitalised countries is represented by an insufficient national internet infrastructure.
Ethical and legal barriers for the adoption of cloud and edge systems	<ul style="list-style-type: none"> • Data handling (need for anonymisation of data, personal data protection). • Commercial in confidence. • The transmission of sensitive information to the cloud. • Possible unauthorised access. • Data corruption. • Infrastructure failure. • Trustworthiness, transparency to the cloud user functionalities is pivotal. • GDPR compliance and compliance with National legal legislation might be different with the cloud provider. Operational applications and services will include personal information that needs to be anonymised. • In case of Teaching Factories where pilot facilities are provided in kind by external partners, IPR, confidentiality and privacy issue might arise when gathering data. In real industrial scenarios (e.g., manufacturing companies) internal weak data management policies, security measures, security certification might represent further barriers to data management. • The aspects of integrity and Company Use security are technologically challenging and might hurt significantly the interests of individuals and of larger parts of society if no appropriate technology is implemented in new production systems. • Although some extra EU Countries are GDPR compliant (e.g., Norway, U.K., Iceland, Lichtenstein, Switzerland), others are not, and legal issues could arise if the cloud servers are located there. Furthermore, for companies that evaluate the “ethical score” of their subcontractors in

	terms of their ESG policy, this evaluation would be more difficult for cloud services, as these may be anywhere in the world.
Potential ethical and legal solutions to the aforementioned barriers	<ul style="list-style-type: none"> • Anonymisation and pseudonymisation policies. • GDPR compliance: definition of data governance structures, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights concerning access to and processing of the data as indicated in Eu Data Act. • Privacy Impact Assessment and development of a single data EU market. • Pooling European data in key sectors, with common and interoperable data spaces. • It has been noticed that is often easier to achieve systems security when situated on the cloud, and also to implement Disaster Recovery Systems and Redundancy Systems, as well as scalable and virtually unlimited retention period of records.
Legal differences in the adoption of cloud or on premises systems	<ul style="list-style-type: none"> • The majority of the partners underlined that the main difference is actually due to the fact that the current legal scenario does not regulate the data sovereignty issues that act as barriers to adopt external cloud-based solutions. There are no legal proofs that the data are not going to be transferred or re-used outside the country of origin, with or without the consent of the data owner. • A partner, instead, reported not to have seen any difference between on-premises and cloud system, explaining that the main decision-making factor is the economic evaluation time after time. • Another partner pointed out that the apparent absence of differences is the reason behind the fact that Teaching Factories are important entities to foster the testing and adoption of cloud/edge and AI systems: by simulating preindustrial manufacturing scenario, they provide structured context for experimentation, enable where appropriate in a real-world environment the testing of cloud/edge AI system overcoming regulatory barriers and company internal policies (e.g., regulatory sandboxes). • A partner believes to be legally easier to adopt cloud-based systems, as some of the security obligations for some specific sectors (e.g., ports) originate from the National Computer Security Incident Response Team regulations.
Current regulation in the adoption of cloud/edge and AI systems	<ul style="list-style-type: none"> • The majority of the partners was aware of the fact that different geographies have different regulations and different approaches to the topic “Privacy vs. Innovation”. • From a European Union perspective, those regulations have been reported: European Data Strategy; Data Governance Act; EU Cybersecurity Act; Proposal for a Regulation laying down harmonised rules on artificial intelligence; Coordinated Plan on Artificial Intelligence (2021). • From a national perspective, for Italy, the National Strategy on AI and Italian Plan on Industry 4.0 have been mentioned. • From a national perspective, for Romania, Romanian Cloud GEO has been mentioned. • On company level, a partner reported to have decided only to use cloud servers within the EU because of the lack of regulation outside the EU borders.

Presence of tax incentives for the	<ul style="list-style-type: none"> • In Latvia, it has been reported the presence of an Industry 4.0 initiative to incentivise the adoption of EtC solutions.
---	--

adoption of edge to cloud systems in EU Countries	<ul style="list-style-type: none"> • In Italy, it has been reported the general aim of national tax credit at supporting the adoption of 4.0 solutions for manufacturing companies including edge to cloud system. In addition, R&D grants support the adoption and development of Industrial 4.0 technologies. • In Romania, there are limited tax incentives that apply for IT and RND projects. • In Cyprus and Germany, no noteworthy incentive has been reported.
Environmental positive impacts related to the adoption of edge to cloud systems	<ul style="list-style-type: none"> • Better work processes enabled by edge to cloud systems (and in general, by new levels of automation) lead to more optimised usage of input materials and lower environmental impact. The partners agreed on the fact that the introduction of a new cutting-edge technology could improve the whole system making it more efficient. • Cloud is more environmentally friendly as unused processing power is usually diverted to other instances, whereas on-premises is exclusively utilised exclusively for internal processes. • The limitation of physical records (paper, and so on) and the use of hardware resources on site (servers, server rooms, climate control, and so on) reduce the Carbon Footprint. • The ability for staff, encouraged and facilitated by the solution, to remote working reduces the need to commute.