Grant Agreement No.: 101070473
Call: HORIZON-CL4-2021-DATA-01
Topic: HORIZON-CL4-2021-DATA-01-05
Type of action: HORIZON-RIA

D1.2 FLUIDOS Governance updated

Work package	WP 1
Task	Tasks 1.1, 1.2, 1.3
Due date	31/08/2024 (extended in consultation with PO)
Submission date	31.08.2024
Deliverable lead	Borderstep Institute
Version	1.0 for submission
Authors	Simon Hinterholzer, Albert Seubers, Nasir Asadov
Reviewers	Fulvio Risso, POLITO James Philpot, DIGITAL SME

Abstract	This document updates D1.1 to include the latest work in the tasks 1.1, 1.2 and 1.3 on governance-, sustainability- and business models related to FLUIDOS

Keywords	FLUIDOS, governance, sustainability, business models

Document Revision History

Version	Date	Description of change	List of contributor(s)

Disclaimer

The information, documentation, and figures available in this deliverable are written by the "Flexible, scaLable and secUre decentralIzeD Operating System" (FLUIDOS) project's consortium under EC grant agreement 101070473 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

Copyright notice

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable: R	Data Management Plan	
Dissemination Level		
PU	Public, fully open, e.g. web	✓

** R: Document, report (excluding the periodic and final reports)*

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patent filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

Funded by Horizon Europe Framework Programme of the European Union

TABLE OF CONTENTS

1. INTRODUCTION

FLUIDOS (Flexible, scaLable, secUre, and decentralIseD Operating System), is an EU funded research project, that aims to leverage the enormous, unused processing capacity at the edge, scattered across heterogeneous edge devices that struggle to integrate with each other and to coherently form a seamless computing continuum. FLUIDOS is structured in 10 work packages, with this document presenting the results of WP1, to update the deliverable D1.1. Work package 1 is divided into three sub-work packages (tasks):

1.1 Infrastructure and data governance models for dynamic peer-to-peer computing ecosystems
1.2 Environmentally sustainable models for decentralised computing
1.3 Business models for fluid computing

These tasks are reflected in the structure (chapters) of this document and are only supplemented by policy recommendations in Chapter 5. It is important to mention that there are many cross-references to the other work packages in FLUIDOS, which are often described in more detail there, while their findings and conclusions are related to each other at a higher level here. Task 1.1 builds heavily on WP 2 reference architecture (including the REAR protocol), WP 3 modular/extensible node, WP4 Intent-based decentralised FluiDOS continuum, and WP 5 security / zero trust. Task 1.2 builds on the findings from WP 6 on cost-effective and energy-aware infrastructure. In Task 1.3 there are some links to WP 7 with the use cases and the market analysis.

This document is based on inputs derived from multiple sources, including desk research and literature review. It also incorporates insights gained through exchanges with other EUCloudEdgeIoT projects and interactions with various stakeholders. Additionally, a detailed survey was conducted to explore the roles and governance within a Cloud-Edge-IoT ecosystem, specifically targeting stakeholders who were unable to participate in the workshop. Furthermore, the document includes findings from a stakeholder workshop held on 27.06.2024, which brought together operators from cloud infrastructures and network providers. This multifaceted approach ensures a well-rounded and robust analysis, providing a solid foundation for the conclusions and recommendations presented herein.

2. Infrastructure and governance models

In Deliverable 1.1 elements of discussion for the definition of a Governance Model were provided, which are formalised in this updated version Deliverable 1.2, following interaction with external and internal stakeholders and progress on the technical development of the FLUIDOS node and the REAR protocol.

FLUIDOS is proposing an overarching operating system that manages and orchestrates distributed edge computing resources, enabling seamless integration, efficient resource allocation and unified control access across heterogeneous devices and platforms. It provides a cohesive layer to handle computing in cloud-edge-iot continuum based subscription and publication of available resources through a dedicated protocol.

Sharing resources in an ecosystem composed of multi-faceted stakeholders can work only if this provides balanced benefits to all the stakeholders and brings guarantees on data security and data privacy. In the case of FLUIDOS, there is an opportunity for infrastructure owners to share their resources (especially at the "edge", i.e. more localised than traditional cloud providers) for the benefit of potential service providers that today do not have access to such resources unless they procure the required infrastructure.

To act as the operator between demand and supply of compute resources, certain conditions must be met and implementing a comprehensive governance model is crucial for maintaining system integrity, security and performance while fostering user satisfaction. The REAR protocol enables potential customers to know what is available in the provider clusters, hence enabling a true, dynamic market. In addition, it aims at automating (and hiding) the technical steps that enable the customer to connect and consume the resources/services agreed in the negotiation phase, such as setting up a peering.

In this peering process, the listed items are taken into account:

Access control	define and enforce policies for who can access and use resources.
Resource allocation	guidelines for distributed compute resources among different users and applications
Data Protection	Implement measures for data encryption, secure data transmission and storage
Identity management	Authentication and authorisation mechanisms to ensure authorised use only
Standardisation	To ensure sharing compute resources standards need to be defined and implemented
Data Life cycle	How is data collected and processed
Data integrity	Data accuracy and consistency across different entities.
Legal Compliance	Adherence to international and national laws and regulations

This opportunity for infrastructure owners is also a chance to foster a more sustainable usage of IT resources, given that in many cases owners of private IT infrastructure are not fully exploiting them, thus leading to a not very environmentally friendly ratio between energy consumption and computational throughput. The sustainability aspect also turns to extending the lifetime of compute resources in some cases as this is not only an economical benefit.
Using shared resources instead of owned resources, also provides the options to select resources that are more sustainable or powerful then the resources in the owned domain.

The balance of sharing and use of shared resources becomes more important with the extension of IoT and edge devices with limited compute resources. These need to rely on the availability of resources they can access on a basis of shared or contracted use.

Economic impact and competitiveness for Europe are to be considered by increasing edge compute power for shared use supported by a meta operating system like FLUIDOS.

The infrastructure governance is built in the process that FLUIDOS follows to share resources, with details described in D3.1. The FLUIDOS Node is adopted to manage the workflow, and control fair and trusted use. Peering candidates abusing fair and trusted use can be purged from the Discovery list.

Besides trust between FLUIDOS nodes and super-nodes, it is important to establish security between the FLUIDOS Edge infrastructure and the edge/IoT devices. To do that, it is fundamental to support trusted computing devices, operating systems, edge microservice communications and networking. Trust between FLUIDOS Edge components and edge devices can be achieved by adopting protocols such as TLS which can offer mutual authentication and confidentiality (data encryption) between two communicating parties. Another important aspect related to security on the edge is to provide secure storage for sensitive data such as keys, certificates, and credentials. One of the main security goals of FLUIDOS project is to ensure the secure execution of workloads, requests and response messages across the FLUIDOS ecosystem, leveraging the capabilities of Trusted Execution Environments (TEEs) and considering different possible implementations. The work in this area mainly focused on how to investigate the authenticity and integrity of an environment by means of Remote Attestation (RA) mechanisms.

Fair use policy implementation ensures the proper use of the promoted computing and storage resources. Since edge resources operate closer to the data source the use of these resources is more interesting for FLUIDOS users. The REAR protocol manager works from the permissible use of promoted resources, particularly in scenarios where data is cached, processed, or analysed at the edge. Additionally, fair use accounts for the potential limitations in monitoring and controlling data at the edge, emphasising the importance of transparency and accountability among users. This includes providing clear instructions for users and developers implementing technical measures to detect and prevent misuse.

3. Environmentally sustainable models

3.1. Energy data disclosure

In our research following policies and incentives to encourage cloud service providers to disclose detailed energy consumption data to their users were identified:

1. Mandatory Reporting Requirements: Implement strict regulatory requirements that compel cloud providers to report energy consumption and carbon emissions regularly (Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on Energy Efficiency and Amending Regulation (EU) 2023/955 (Recast), 2023).
2. Tax Incentives: Offer tax rebates or reductions for cloud providers that demonstrate energy efficiency and transparency in their operations.
3. Customer Demand Initiatives: Encourage collective customer demands or campaigns urging providers to disclose energy usage data.
4. Eco-Certification Programs: Develop and promote certification programs that recognize and reward transparency and efficiency in energy use (Peter Judge, 2023).
5. Partnerships with Environmental Organizations: Foster partnerships between cloud providers and environmental NGOs to create standards and best practices for energy disclosure.
6. Transparent Billing Statements: Include detailed energy consumption data in user billing statements, similar to utility bills.
7. Competitive Differentiation: Encourage cloud providers to use energy transparency as a marketing tool to attract environmentally conscious customers (Rahkonen & Dietrich, 2023).

During a workshop organized by the FLUIDOS consortium, which included industry representatives, this question was presented in a quiz format (see Figure 1).

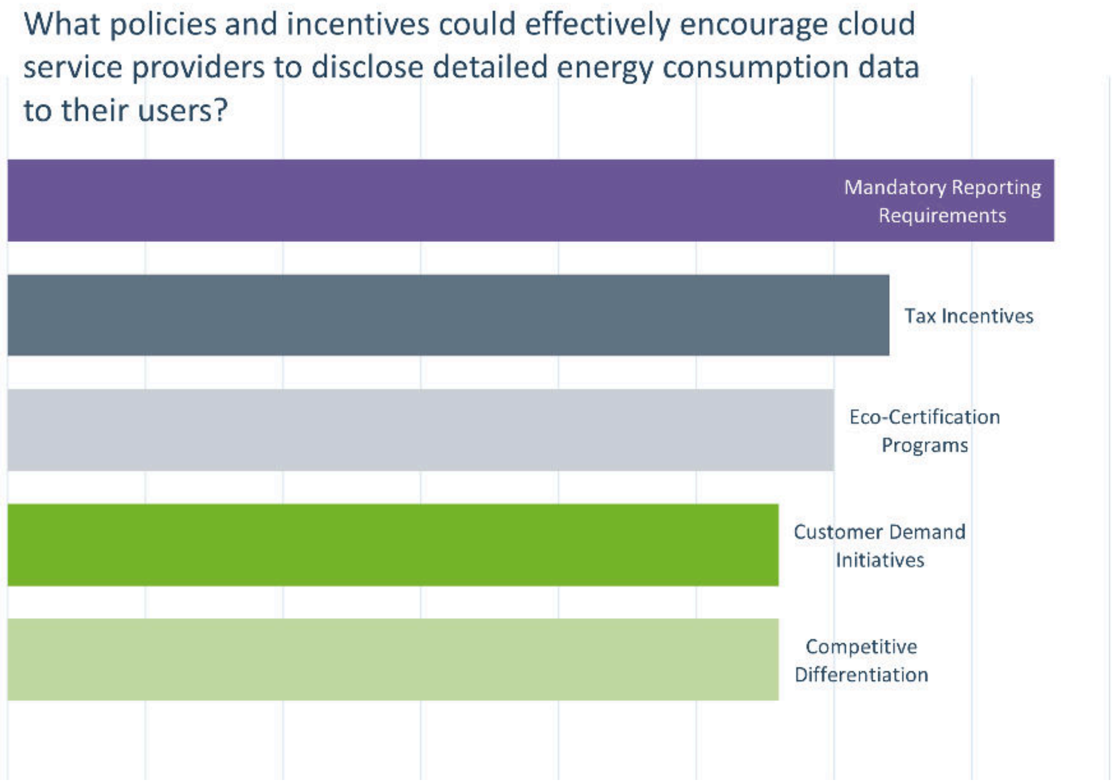

Figure 1: Results from the live survey on data disclosure

The questionnaire results indicate that the most effective policy to encourage cloud service providers to disclose detailed energy consumption data is implementing mandatory reporting requirements, followed closely by offering tax incentives. Eco-certification programs also garnered significant support, highlighting the value of recognizing and rewarding transparency and efficiency in energy use. Customer demand initiatives were seen as another effective strategy, reflecting the belief that consumer pressure can drive industry change. Lastly, while competitive differentiation was acknowledged as beneficial, it received the least support compared to the other measures. Overall, regulatory and incentive-based approaches are preferred, with mandatory reporting being the most favored.

3.2. Circular economy models

Following models to ensure environmentally sustainable lifecycle of edge computing infrastructures were identified in our research:

1. Design for Disassembly: Encourage the design of edge devices that can be easily disassembled for recycling or refurbishing (Formentini & Ramanujan, 2023).
2. Extended Producer Responsibility: Implement policies requiring manufacturers to manage the disposal and recycling of their products at end of life (The Role of Extended Producer Responsibility in E-Waste Management and Recycling, 2024).
3. Subscription and Leasing Models: Promote business models where edge devices are leased to customers and returned to the provider for upgrading or recycling.
4. Modular Design: Support the development of modular edge devices that can be easily upgraded with minimal waste.

5. Use of Recycled Materials: Incentivize the use of recycled or biodegradable materials in the manufacturing of edge devices.
6. Life Cycle Assessment Requirements: Require companies to conduct and publish life cycle assessments for products to inform sustainable development (Kloepffer, 2008).
7. Green Procurement Standards: Adopt procurement policies that prioritize buying edge technologies that adhere to environmental sustainability standards.

Similar to Section 3.1, this question was provided to workshop participants in quiz format (see Figure 2).

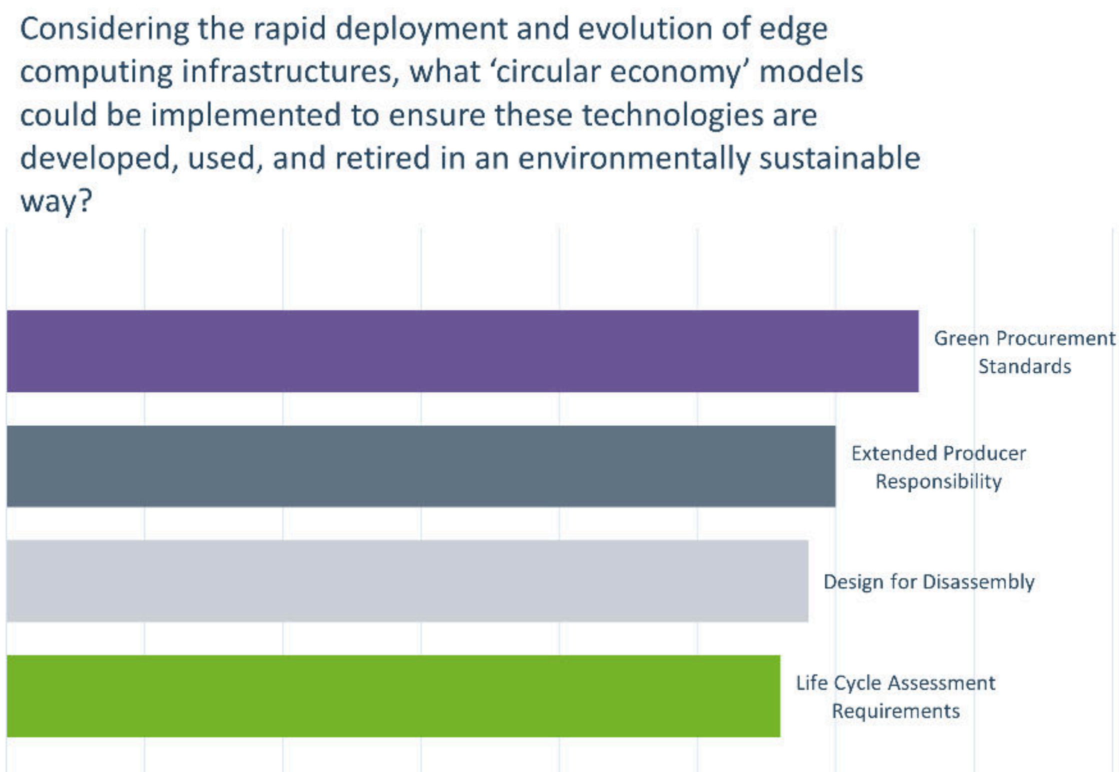

Figure 2: Results from the live survey on 'circular economy' models

The questionnaire results indicate that the most favoured 'circular economy' model for ensuring the environmental sustainability of edge computing technologies is the implementation of green procurement standards. This is followed by extended producer responsibility, which holds producers accountable for the entire lifecycle of their products. The design for disassembly also received significant support, emphasizing the need for products that are easy to disassemble for recycling or reuse. Life cycle assessment requirements, while still supported, received the least endorsement among the listed measures. Overall, there is a strong preference for standards and responsibilities that promote sustainability, with green procurement standards being the most favoured approach.

4. Business models for fluid computing

After conducting the first internal business model workshop in year 1, it became clear that further in-depth analysis and the collection of external perspectives were necessary for various questions regarding the business models in FLUIDOS, and the online expert survey described above and the workshop with external partners on 27.06.2024 were used for this purpose.

While FLUIDOS is ideally suited to providing resources for other people/services free of charge within your own organization (which is already demonstrated in the internal project use cases in WP7), the following section looks at business models in which compute resources are provided in the network for a fee; as illustrated by the simplified Figure 3 below.

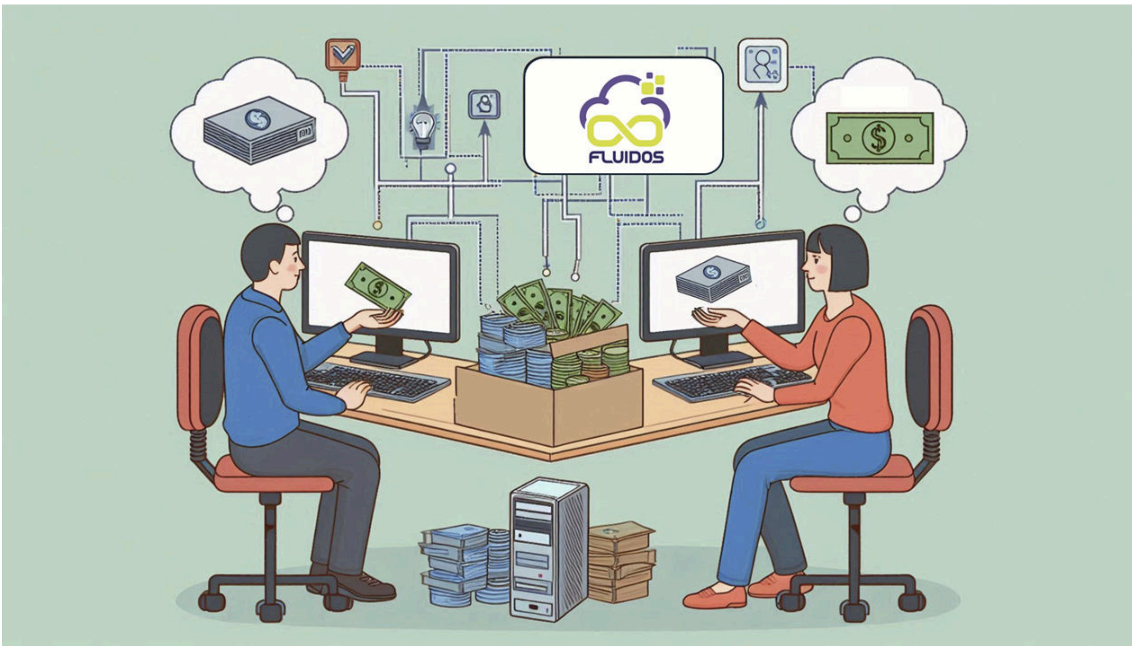

Figure 3: Sharing resources in the network with cash in return; image created with AI of Microsoft Designer and some manual edits

Such edge infrastructures can be deployed in very different ways (see fig. 4). Some might be provided on the user-edge (also known as far edge). A considerable number of networked end devices are equipped with surplus capacity, including CPU cores, RAM, and storage, which can be made available on the network. Similarly, there is often considerable capacity on the user side, as in the case of production facilities for Industry 4.0. This capacity can be flexibly allocated to different external or internal applications, and may comprise a range of hardware, from industrial computers to desktop PCs or powerful servers.

Service providers may also offer capacities at their network edge, which are marketed as follows:

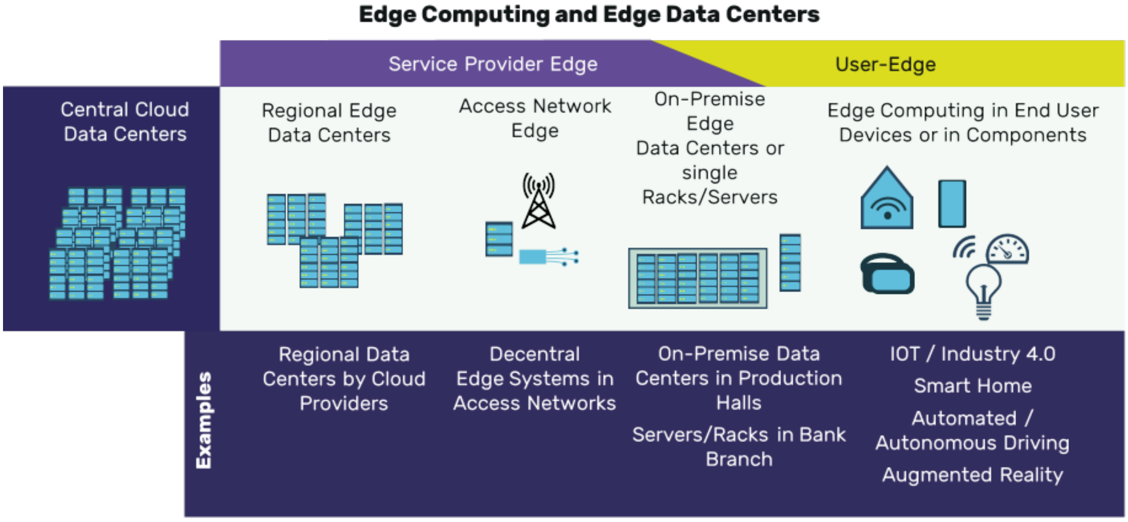

Figure 4: The different operators, locations and types of physical edge resources

In most of the cases, the essential characteristics of cloud computing as defined by NIST[1] also apply to the characteristics of IaaS services at the edge:

> *"On-demand self-service*. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
>
> *Broad network access*. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
>
> Resource pooling. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data centre). Examples of resources include storage, processing, memory, and network bandwidth.
>
> *Rapid elasticity.* Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.
>
> *Measured service*. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be

[1] Definition of the characteristics of cloud computing, Mell & Grance, September 2011.

monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

One of the few differences between FLUIDOS and conventional cloud computing is that more information about the hardware, location and energy mix is made transparent. Only by this transparency, the very heterogeneous infrastructures at the edge be utilized sensibly, and the workloads distributed according to economic and ecological criteria. The elasticity to scale may be limited for individual domains/clusters, but for the entire Continuum, which includes also the cloud, the elasticity is virtually unlimited.
In the following, the results of the workshop in FLUIDOS are analysed in this context and some core elements of the business model behind the sale of edge resources (value proposition, cost-, revenue- and market model) are analysed.

4.1. Evaluation of the workshop and the survey regarding business models

At the core of a business model is a product and its value proposition for the customer. For FLUIDOS, the product is quite similar to existing products in the field of cloud computing, i.e. primarily virtualized services to obtain storage or computing power, or the "X-as-a-service" products based on them. FLUIDOS is not an alternative to the central cloud, but rather aims to tap into previously underutilized resources at the edge and incorporate them into a cloud-to-edge continuum. The aim is not to bring computing power exclusively from the cloud to the edge, but rather to harness the advantages of both worlds through greater transparency and to make the shifting of workloads simpler, and also dynamic (e.g. depending on price or environmental impact) and transparent (driven by the FLUIDOS orchestrator instead of being manually set by the infrastructure operator or the service provider).

When asked what the biggest advantage of such fully virtualized resources at the edge could be, many participants saw a high value in the better latency (see figure 5). This latency depends primarily on the physical proximity and even more on the number of network hops from the end user/application device to the (central) data processing. If an application can access (network topologically) very close data processing capacity, it is more likely that this can be reached with low latency.
Local "on-premise" computing is also seen as a major advantage. This can have completely different motivations; for example, one could argue economically that by tapping into local (in-house) capacities, it may be possible to save money that does not have to be spent on cloud resources. On the other hand, it can also be motivated by data protection reasons that you prefer to store and process certain data on local systems. Other points mentioned in the comments were the privacy of the user data/data ownership as well as security, which may be related to this.

Another major value lies in the use of previously underutilized resources, which were ranked third and fourth by the participants, together with *Independence* from the cloud. This is due to the fact that the performance of data processing has increased massively in recent years, particularly on end devices, but often also on servers, although this maximum performance is only very rarely or not at all required. FLUIDOS aims to tap into this unused capacity.

Figure 5: Results from the live survey on the value of a fluidified computing environment at the edge

When considering the benefits of a service, it is very important to understand which customer will use it. Especially with the variety of potential cloud edge services in an IaaS business model, there are many potential applications and customers who operate these applications. To determine which customers are most relevant for highly distributed computing resources, we asked this question in the workshop, see fig. 6.

Unsurprisingly, IoT service providers were ranked highest. An IoT service provider offers platforms and tools to connect, manage, and analyse data from a multitude of IoT devices. They enable seamless communication between devices and ensure efficient data processing, often facilitating real-time analytics and decision-making. Distributed compute resources at the edge are crucial for these providers because they reduce latency, enhance response times, and allow for localized data processing, improving the overall performance and reliability of IoT applications.

Network operators (esp. mobile networks) were also ranked very high. A network operator is a company that provides communication services to various device users by owning or controlling the necessary infrastructure and spectrum licences in mobile networks. They ensure connectivity, manage network traffic, and offer services such as voice, text, and data. Distributed compute resources at the edge are crucial for network operators because they reduce latency, improve data processing speeds, and enhance the quality of service for applications like video streaming, gaming, and IoT, leading to a better user experience and more efficient network management.

IT companies in general and companies with high demand for AI / machine learning capacities are ranked in the middle. For companies with high demand in AI capacities, edge computing can be vital as it allows for efficient processing of large datasets close to the data source, minimizing delays and bandwidth usage, and enabling rapid and more effective deployment of AI models and algorithms (inference). This leads to more responsive and scalable AI-driven solutions. Research institutions and cloud providers were ranked lowest.

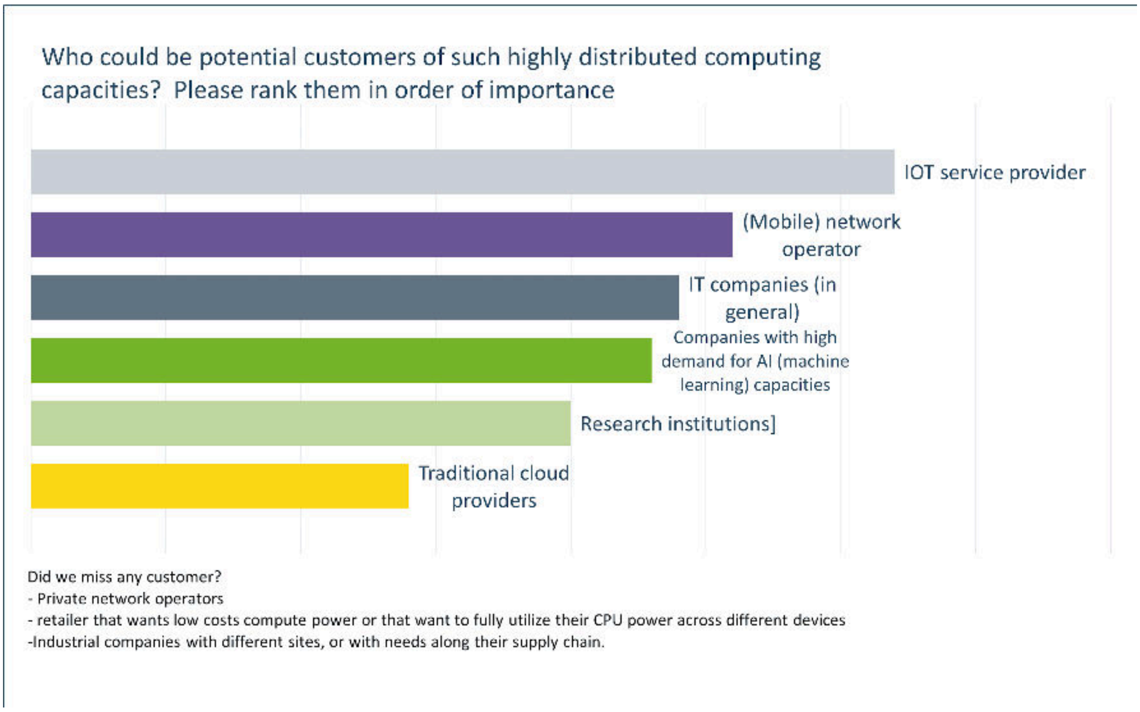

Figure 6: Results from the live survey on the relevance of different customers for highly distributed computing capacities

4.2. Cost and revenue for compute resources at the edge

This section will discuss the cost and revenue part of a business model to provide compute resources at the edge.

Cost model

In an Edge-to-Cloud ecosystem, the proximity of edge computing resources to end users can be leveraged, while utilizing the vast computational and storage capacities of cloud data centres. This hybrid approach aims to optimize performance, latency and resource utilization. The cost and revenue model in such an ecosystem involves several components, including capital expenditure (CAPEX), operational expenditure (OPEX), pricing strategies, and revenue streams. Unused compute and storage resources of end devices can also be considered in the area of (far) edge computing, which can be made available to the network. Another opportunity with FLUIDOS is the dynamic relocation of workloads across the boundaries of the cloud ecosystems of individual providers, in particular to optimize costs and/or carbon emissions.

The Capital Expenditure (CAPEX) involves substantial initial investments:

- Edge Infrastructure Costs include expenses for edge servers, gateways, and networking equipment, alongside deployment costs such as installation, site preparation, and configuration. This also covers edge devices like sensors and IoT peripherals. Some of the physical edge infrastructure can be considered to be available anyway (e.g. in notebooks or smartphones), but are just not fully utilized (i.e., no sharing) without FLUIDOS.
- Cloud Infrastructure Costs cover data centres with servers, storage systems, and networking hardware, including high-speed interconnects and backbone networks. Investments in redundancy and backup systems for failover capabilities and disaster recovery are also included.

The Operational Expenditure (OPEX) covers ongoing costs:

- Edge Operations encompass maintenance costs for edge devices, firmware updates, repairs, and energy consumption. Networking costs for data transfer between edge and cloud, and salaries for on-site technicians and network engineers, are also part of this category. In some cases, devices would run anyway and the existing network capacity from WiFi and/or mobile networks are sufficient and available for free.
- Cloud Operations include power and cooling costs for data centres, maintenance of servers and networking equipment, salaries for operational staff, and software licences for necessary applications and virtualization.

Other costs that need to be accounted for include:

- Security investments in cybersecurity measures for both edge and cloud components.
- Compliance costs associated with meeting regulatory and compliance requirements.
- Scalability expenses for scaling infrastructure to meet increasing demand.

A key aspect of the edge computing model is the dynamic utilization of end user device resources. Unused compute and storage resources of devices can be dynamically made available to the network, significantly reducing infrastructure costs and improving resource utilization by leveraging existing hardware.

Pricing and payment

A pricing and payment model for resource allocation and scheduling at the edge could incorporate several key elements to ensure efficiency, fairness, and incentivization. The characteristics of payment models derived from cloud computing are of particular relevance to the pricing and trading of compute resources in the edge area:

1. Dynamic pricing: Prices fluctuate based on supply and demand, which logically leads to higher prices during peak usage times and lower prices during off-peak hours
2. Resource-based pricing:
    - Different rates depending on CPU, memory and storage. Typically, per CPU core (vCPU) or per GB of memory
    - Pricing tiers based on performance levels (e.g., standard vs. high-performance)
3. Location-based pricing:
    - Varying rates depending on the geographical location of edge resources
    - Higher prices for high-demand areas or regions with limited infrastructure
4. Quality of Service (QoS) pricing:

- Premium pricing for guaranteed low-latency or high-availability services
- Basic pricing for best-effort services

5. Time-based pricing:
    - Rates based on duration of resource usage
    - Discounts for long-term commitments or reservations
6. Auction-based allocation:
    - Real-time bidding for available resources
    - Spot pricing for unused capacity
7. Tokenization and micropayments:
    - Use of blockchain or distributed ledger technology for secure, rapid transactions
    - Ability to make small, frequent payments for granular resource usage
8. Incentive mechanisms:
    - Rewards for end-users who share their device resources
    - Bonuses for consistent availability or high-quality service provision
9. Federated pricing models:
    - Collaboration between different edge providers to offer unified pricing across networks
    - Revenue sharing agreements for cross-network resource usage
10. Usage-based billing:
    - Pay-as-you-go model for flexible resource consumption
    - Ability to set usage limits or budgets to control costs
11. Subscription-based options:
    - Flat-rate plans for predictable workloads
    - Tiered subscriptions with different resource allowances
12. Multi-factor pricing:
    - Combining multiple pricing elements (e.g., base rate + usage + QoS premium)
    - Customizable pricing plans based on specific user needs

This framework aims to balance the needs of resource providers, end-users, and network operators while optimizing resource allocation and encouraging efficient usage of edge computing capabilities.

The location-based pricing has a special character in edge computing, as it can not be reduced to a manageable number of availability zones as in cloud computing, but will represent many more physical locations (and logical locations in networks).

Large cloud computing providers already offer capacities with different prices depending on location (and also other factors like time). The following screenshots show the prices of "general purpose" instance types in two different locations/availability zones (Zurich and Frankfurt) on 26th of July 2024 between 8am and 9am (see fig. 7 and fig. 8):

Instance name	On-Demand hourly rate	vCPU	Memory	Storage	Network performance
t4g.nano	$0.0053	2	0.5 GiB	EBS Only	Up to 5 Gigabit
t4g.micro	$0.0106	2	1 GiB	EBS Only	Up to 5 Gigabit
t4g.small	$0.0211	2	2 GiB	EBS Only	Up to 5 Gigabit
t4g.medium	$0.0422	2	4 GiB	EBS Only	Up to 5 Gigabit
t4g.large	$0.0845	2	8 GiB	EBS Only	Up to 5 Gigabit
t4g.xlarge	$0.169	4	16 GiB	EBS Only	Up to 5 Gigabit
t4g.2xlarge	$0.3379	8	32 GiB	EBS Only	Up to 5 Gigabit

Figure 7: Pricing of example cloud RESOURCES in ZURICH

Instance name	On-Demand hourly rate	vCPU	Memory	Storage	Network performance
t4g.nano	$0.0048	2	0.5 GiB	EBS Only	Up to 5 Gigabit
t4g.micro	$0.0096	2	1 GiB	EBS Only	Up to 5 Gigabit
t4g.small	$0.0192	2	2 GiB	EBS Only	Up to 5 Gigabit
t4g.medium	$0.0384	2	4 GiB	EBS Only	Up to 5 Gigabit
t4g.large	$0.0768	2	8 GiB	EBS Only	Up to 5 Gigabit
t4g.xlarge	$0.1536	4	16 GiB	EBS Only	Up to 5 Gigabit
t4g.2xlarge	$0.3072	8	32 GiB	EBS Only	Up to 5 Gigabit

Figure 8: Pricing of example cloud RESOURCES in FRANKFURT

For the selected instance types, it can be stated that prices in Zurich are approx. 9% higher than in Frankfurt. Even if the instance types at the edge are in many cases very different from those in cloud data centres (performance, availability, bandwidth, etc.), one can ask whether they are in competition with each other. For many applications or parts of applications (microservices), the advantages of the cloud certainly outweigh the disadvantages. However, local provision at the edge has latency advantages and may be more appropriate to guarantee requested data governance rules, which in some cases represent a unique selling point and therefore also allows higher prices. In addition, especially for IoT/far edge, provisioning could take place on devices that are running anyway, which is why the marginal costs of provisioning are almost negligible as long as the actual use of the end device/IoT device is not affected.

At the workshop, most participants (67%) were very clear that a pay-per-use payment model makes the most sense for such edge payment processes. Only 5% voted for subscription-based payment models and the rest (28%) thought hybrid models were the most promising (see fig. 9).

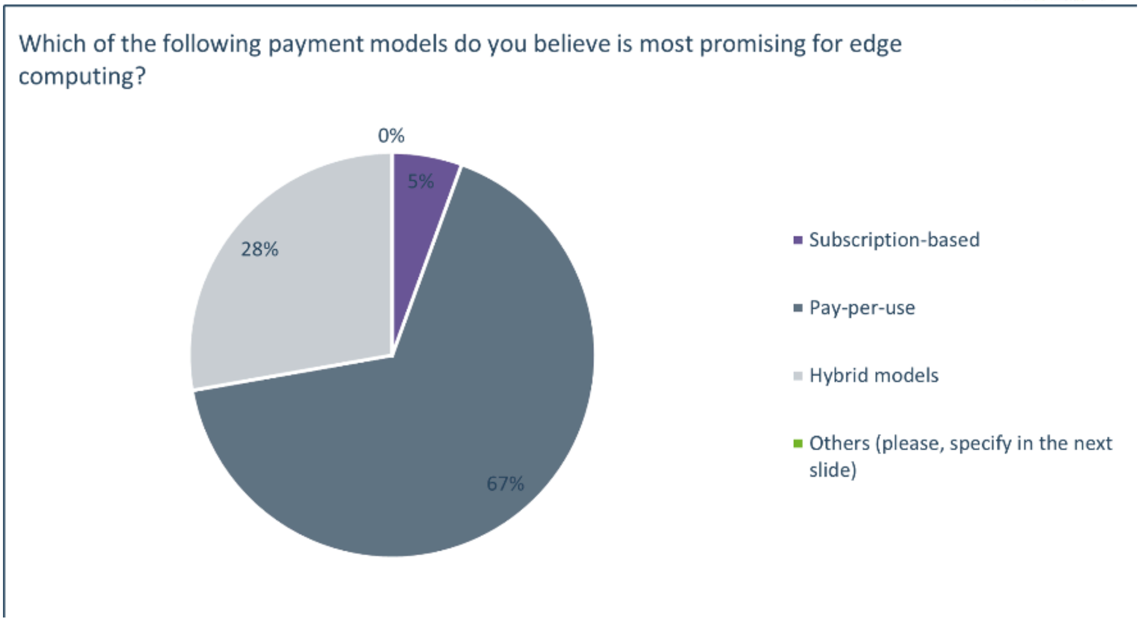

FIGURE 9: RESULTS FROM THE LIVE SURVEY ON PAYMENT MODELS

Finally, it is worth considering that in several cases **payment models are not needed**, particularly when all the involved resources belong to the same administrative domain. In fact, an interesting deployment model for the FLUIDOS metaOS is when all the resources belong to the same entity (e.g., the same company), which may not be interested to activate revenue-based business models within the organization itself, provided that the FLUIDOS approach guarantees better resource utilization, hence bringing substantial cost savings.

Competitiveness of edge infrastructures

Since approximately 2017, a number of online journals and articles have published forecasts predicting that edge computing will replace a significant proportion of cloud computing,or even replace them almost completely (e.g. Baker, 2023; Bittman, 2017; Kleyman, 2018; Raza, 2023). In recent years, however, a number of articles have also predicted the clear coexistence or supplementation of cloud computing with edge infrastructures (Pavel Despot, 2024; Robinson, 2022).

Multiple users expect that sharing CPU or storage resources across different domains will reduce their cost (see fig. 10). Some, at least, mention that this will only happen if the installed base of hardware can be reduced, which makes sense as IT is one of the major CAPEX cost factors cost of data centres (Luiz André Barroso et al., 2018).

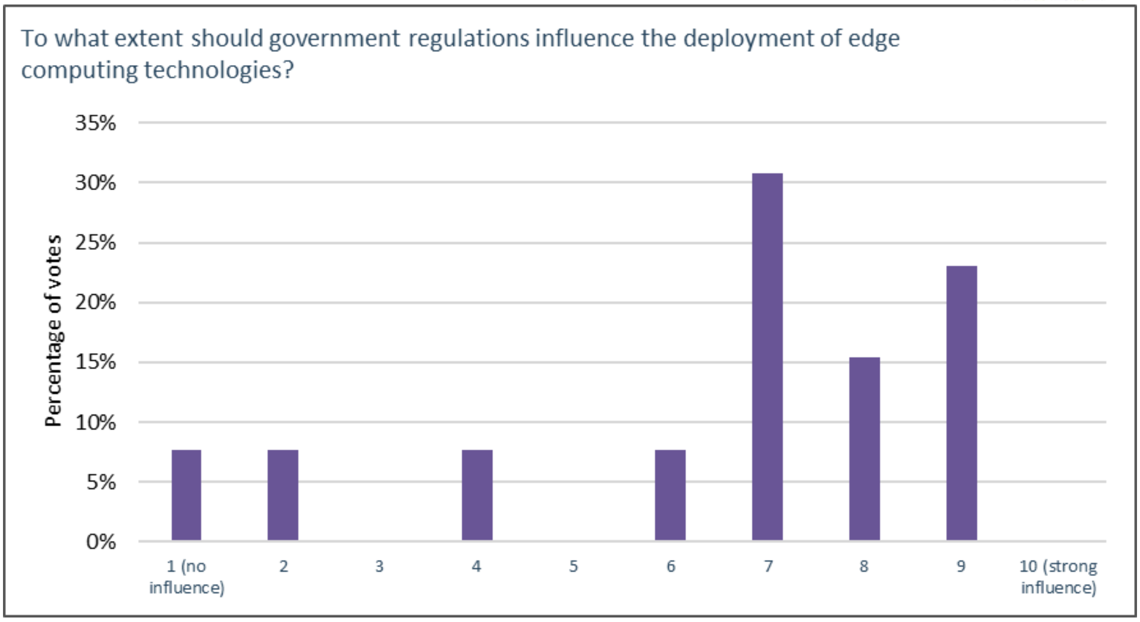

Figure 10: Results from the live survey on cost reduction by sharing resources across different domains

It is yet to be determined what form competition between different edge cloud continuums (i.e. other MetaOS environments) will take in the future, given that the markets are not yet sufficiently mature to support such a scenario. The Deliverable D7.1 Market Analysis examines analogous (i.e. potentially competing) projects (if one can even apply such a description to open source projects).

Revenue model

The extent to which such low-latency computing power will be required, and the price users are prepared to pay for it, will depend heavily on the development of specific applications such as autonomous driving, augmented reality, Industry 4.0, online gaming, etc. Their future development and computing requirements at the edge are currently barely predictable.

In principle, revenue would result from the usage-based proceeds from the sale of computing resources, plus any subscription fees if a hybrid payment model is offered/used. If an operator of the physical edge infrastructure itself offers higher service levels or direct applications (FaaS, SaaS, etc.), even higher revenues can be generated, but this is not in the scope of this report.

The monthly revenue from subscriptions would be calculated relatively simply by multiplying the monthly fee by the number of subscribers.
The pay-per-use revenue is somewhat more complex. While costs for storage services can be measured relatively easily in terms of data volume, availability and security (data redundancy), compute services (CPU, memory) are somewhat more difficult to measure. In the cloud sector, cores of large CPUs are sold as vCPUs, with the generation of the CPU typically determining how a vCPU is priced. However, due to the high heterogeneity of edge devices, it is a challenge to compare resources and price them fairly. In this case, a standardised benchmark could possibly provide representative information on performance. Such a benchmark might have to be carried out automatically by an external 'neural' actor (e.g. at startup/login). Such a benchmark could be similar to the 'SPEC Cloud® IaaS 2018' benchmark (Standard Performance Evaluation Corporation, 2024), which tests multiple performance indicators in cloud environments.

5. Policy recommendations

During the workshop, the participants were also asked to what extent member states should influence the provision of edge computing technologies. The different views are widely spread. However, there was a relatively strong majority in favour of strong influence, although no one selected the extreme value (10), see fig. 11.

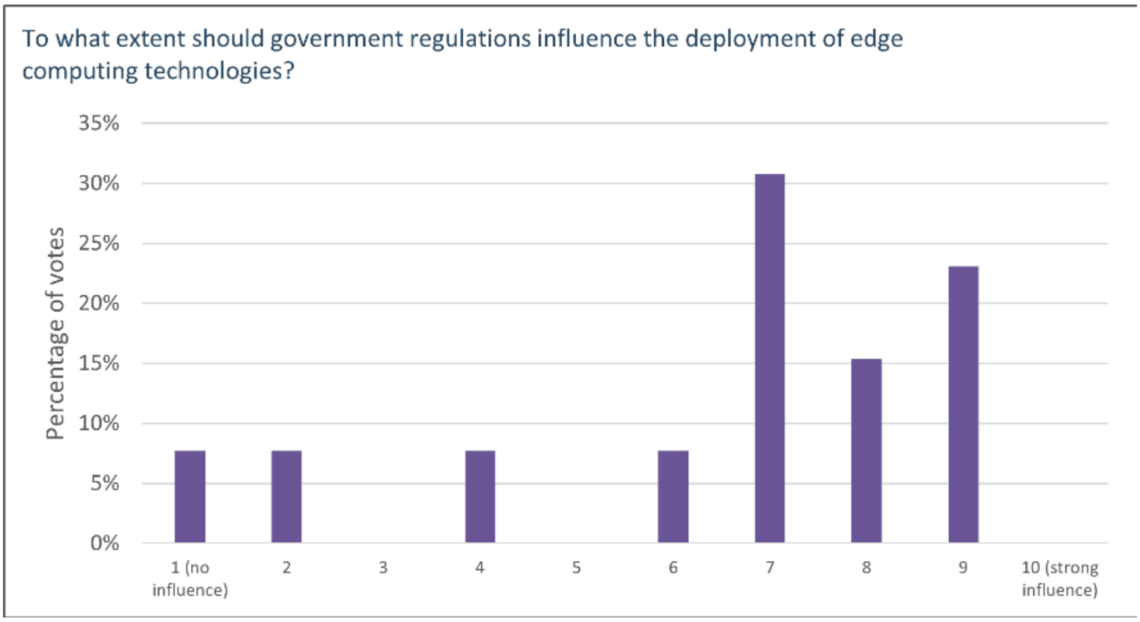

Figure11: Results from the live survey on government regulations to influence the deployment of edge computing technologies

From the perspective of the European Union, there are multiple ways, to support such distributed computing:

Policy, Regulation, and Legal Framework: To support open source based distributed (edge-)computing, governments within the EU can develop comprehensive policies aligned with the Digital Single Market strategy and the European Green Deal. These policies would mandate or incentivize the use of open source solutions in the public sector, mirroring the EU's Open Source Software Strategy. A robust legal framework, building upon GDPR principles, would address data privacy, security, liability, and intellectual property rights in distributed systems. In cooperation with ENISA, governments can establish standards for secure and efficient resource sharing, ensuring compliance with the EU Cybersecurity Act for cloud services and IoT devices. In addition, strong support for distributed computing and carbon aware networks would be an important part of the digital decade goals, which also envisage the operation of 10,000 climate-neutral edge nodes (European Commission, 2024a). This approach would create a solid foundation for the widespread adoption of open source cloud-edge computing solutions while maintaining the EU's commitment to digital sovereignty and cybersecurity.

Funding, Infrastructure, and Economic Incentives: The EU and its member states can leverage existing EU funding mechanisms like Horizon Europe to establish more market-ready demonstrators and lighthouses, which go beyond the levels aimed at FLUIDOS and similar projects (aiming at TRL 5).

They can fund pilot projects similar to EU Smart Cities initiatives, demonstrating the feasibility and benefits of these systems. Financial incentives for organizations adopting such technologies would align with the EU's digital transformation goals. Infrastructure investment, utilizing the Connecting Europe Facility, would improve network capabilities to support distributed computing. Establishing government-run edge data centres in strategic locations, with consideration for EU cohesion policy objectives, would further strengthen the ecosystem. Tax incentives for private companies building green edge computing facilities and procurement policies prioritizing open source solutions in government IT would round out a comprehensive economic strategy to drive adoption and innovation in this field.

Education, Awareness, and Collaboration: To build a strong foundation for open source distributed computing, governments can support educational programs aligned with the EU's Digital Education Action Plan. Organizing hackathons and competitions, possibly in connection with the EU Code Week initiative, would foster innovation in the field. Public awareness campaigns highlighting the benefits of resource sharing and its role in achieving EU climate neutrality goals would drive public support. Fostering public-private partnerships, building on the European Partnership model, and encouraging international cooperation on open standards would accelerate development. Supporting community-driven open source projects and collaborating with universities to integrate cloud-edge computing into curricula, e.g. within the EU digital skills and jobs platform (European Commission, 2024b), would ensure a steady flow of skilled professionals and innovative ideas, in line with the EU's commitment to open science and innovation.

Green Energy Integration and Sustainability: Implementing policies that prioritize workload shifting to areas with abundant renewable energy would support the EU's renewable energy targets. Offering incentives for green energy-powered edge computing facilities aligns with the European Green Deal. Developing smart grid technologies to optimize energy use in distributed systems, building on the EU's Smart Grids Task Force work, would further enhance sustainability. Ensuring compatibility with the EU Taxonomy for sustainable activities in the ICT sector and supporting the development of energy-efficient algorithms would contribute to the EU's energy efficiency goals. Moreover, promoting the use of distributed computing for climate modelling and environmental research would support the European Climate Pact, demonstrating the technology's potential to address critical environmental challenges.

Standardization and Interoperability: Governments can lead or participate in the development of international standards for distributed computing, leveraging the EU's influence in global standard-setting bodies. Ensuring interoperability between different open source solutions, in alignment with the EU's Interoperability Framework, would be crucial. Collaborating with ETSI or IEEE to develop open standards for cloud-edge computing and promoting the adoption of common APIs and protocols (e.g. REAR protocol) across the EU would create a cohesive ecosystem. Supporting the development of open-source reference implementations for key standards and ensuring compatibility with existing EU initiatives like GAIA-X would further strengthen the European cloud-edge computing landscape. This approach would facilitate the creation of a robust, interoperable ecosystem that adheres to EU values and principles, fostering innovation while maintaining European digital sovereignty.

Conclusions

The findings of this report highlight the critical developments and recommendations for the FLUIDOS project in the areas of infrastructure and data governance, sustainability, and business models for fluid computing. The progress made in updating governance models, particularly through the interaction with stakeholders and advancements in the FLUIDOS node and REAR protocol, establishes a robust framework for resource sharing within a decentralized computing ecosystem. The governance model proposed by FLUIDOS is designed to ensure system integrity, security, and performance while also fostering user satisfaction by enabling a dynamic and fair marketplace for computing resources.

The emphasis on environmentally sustainable models is another significant outcome of this research. The report identifies several policy instruments and business models aimed at promoting energy efficiency and circular economy practices within the FLUIDOS framework. These models not only address the environmental impact of edge computing but also encourage the design of sustainable edge devices and the responsible management of their lifecycle. The recommendations for mandatory energy data disclosure and the promotion of green procurement standards are particularly noteworthy, as they align with broader EU sustainability goals and set a precedent for responsible IT infrastructure management.

In the realm of business models, the report underscores the importance of developing flexible and transparent pricing strategies that can accommodate the unique characteristics of edge computing. The proposed models emphasize dynamic pricing based on supply and demand, resource-based pricing, and location-based pricing, which reflect the complexity and heterogeneity of edge resources. The integration of incentive mechanisms, such as rewards for resource sharing and federated pricing models, is crucial for fostering a competitive and cooperative market environment.

The report also provides valuable insights into the potential economic impact and competitive advantages of adopting FLUIDOS across Europe. By leveraging underutilized resources at the edge, FLUIDOS not only enhances the efficiency of existing IT infrastructures but also contributes to reducing operational costs and improving sustainability metrics. The recommendations for policy regulation, funding, and standardization provided in this report are aligned with the European Union's strategic goals for digital sovereignty and climate neutrality.

In conclusion, the updated governance models, sustainable practices, and innovative business models presented in this report are critical for the successful deployment and adoption of FLUIDOS. These elements not only address current challenges in the cloud-edge continuum but also position FLUIDOS as a leader in the transition towards more decentralized, sustainable, and economically viable computing infrastructures. The continued collaboration between stakeholders and alignment with EU policies will be essential for realizing the full potential of the FLUIDOS project.

LITERATURE

Baker, P. (2023). *Edge Computing Eats the Cloud? | InformationWeek*. https://www.informationweek.com/it-infrastructure/edge-computing-eats-the-cloud-

Bittman, T. (2017, März 6). *The Edge Will Eat The Cloud*. Gartner Blog Network. https://blogs.gartner.com/thomas_bittman/2017/03/06/the-edge-will-eat-the-cloud/

Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on Energy Efficiency and Amending Regulation (EU) 2023/955 (Recast), EP, CONSIL, 231 OJ L (2023). http://data.europa.eu/eli/dir/2023/1791/oj/eng

European Commission. (2024a). *Europe's digital decade: 2030 targets | European Commission*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

European Commission. (2024b). *Platform | Digital Skills and Jobs Platform*. https://digital-skills-jobs.europa.eu/en/about/digital-skills-and-jobs-platform

Reworx Electronics Recycling. (2024, Juni 16). *The Role of Extended Producer Responsibility in E-Waste Management and Recycling*. Reworx Electronics Recycling. https://www.reworxrecycling.org/epr-ewaste-management-recycling/

Formentini, G., & Ramanujan, D. (2023). EXAMINING THE ROLE AND FUTURE POTENTIAL OF DESIGN FOR DISASSEMBLY METHODS TO SUPPORT CIRCULAR PRODUCT DESIGN. *Proceedings of the Design Society*, *3*, 1735–1744. https://doi.org/10.1017/pds.2023.174

Kleyman, B. (2018, April). *Edge is Eating Away at the Cloud, and The Data Center Will Benefit*. Data Center Frontier. https://www.datacenterfrontier.com/cloud/article/11430257/edge-is-eating-away-at-the-cloud-and-the-data-center-will-benefit

Kloepffer, W. (2008). Life cycle sustainability assessment of products. *The International Journal of Life Cycle Assessment*, *13*(2), Article 2. https://doi.org/10.1065/lca2008.02.376

Luiz André Barroso, Urs Hölzle, Parthasarathy Ranganathan, & Margaret Martonosi. (2018). *The Datacenter as a Computer: Designing Warehouse-Scale Machines, Third Edition*. Morgan & Claypool. http://ieeexplore.ieee.org/document/8519941

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. *US National Institute of Standards and Technology (NIST)*, *53*(6), 50.

Pavel Despot. (2024). *Edge Computing Versus Cloud Computing: Key Similarities and Differences*. Akamai. https://www.akamai.com/blog/edge/edge-computing-versus-cloud-computing-key-similarities-differences

Peter Judge. (2023, September 21). *European Energy Efficiency Directive published, with mandatory data center reporting*. https://www.datacenterdynamics.com/en/news/european-energy-efficiency-directive-published-with-mandatory-data-center-reporting/

Rahkonen, T., & Dietrich, J. (2023). *Server energy efficiency: Five key insights*. Uptime Institute. https://uptimeinstitute.com/resources/research-and-reports/server-energy-efficiency-five-key-insights

Raza, K. (2023). *Eating the internet: The rise of edge computing and a new business network | Edge Industry Review*. https://www.edgeir.com/eating-the-internet-the-rise-of-edge-computing-and-a-new-business-network-20230116

Robinson, D. (2022). *Edge computing set for growth – that is, when we can agree what it is*. https://www.theregister.com/2022/01/17/edge_computing_set_for_growth/

Standard Performance Evaluation Corporation. (2024). *SPEC Cloud® IaaS 2018*. https://www.spec.org/cloud_iaas2018/