Finanziato dall'Unione europea
NextGenerationEU

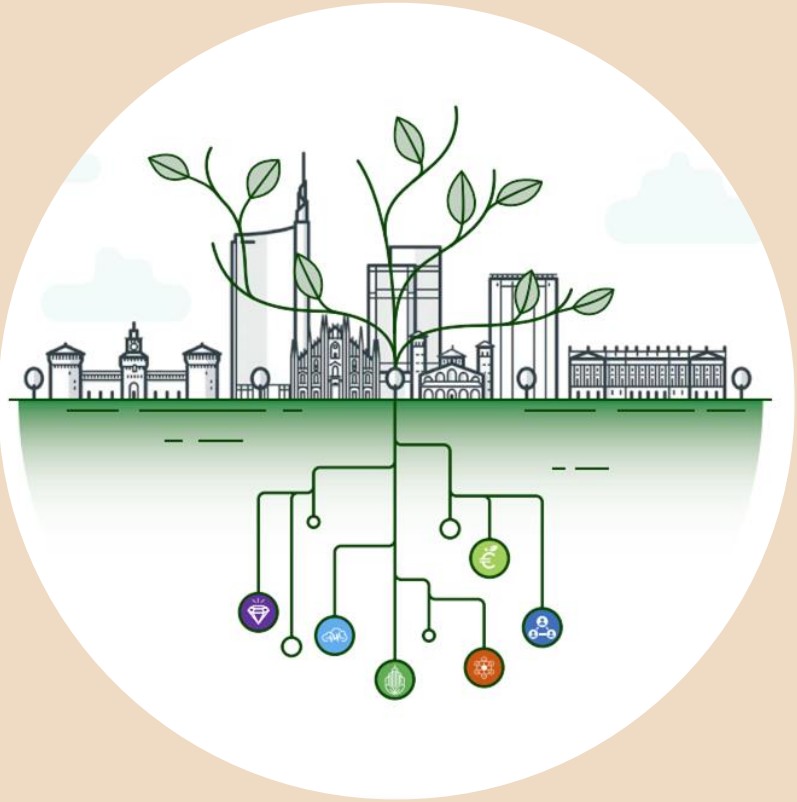
Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

MUSA

PNRR MUSA

Smart Deployment Solution for PNRR MUSA services in Edge-Cloud Continuum



Claudio A. Ardagna
Marco Anisetti

Università degli studi di Milano

27 November 2023

Agenda

- Overview of PNRR Project Multilayered Urban Sustainability Action (MUSA) – Spoke 2
- MUSA digital platform for data analytics pipeline management in the cloud-edge continuum
- Smart Service Deployment Solution in Edge-Cloud Continuum
- Next steps and cascading grant

PNRR MUSA

- **MUSA – Multilayered Urban Sustainability Action**
- PNRR Innovation Ecosystem
- 6 Spoke – 24 Partner - 970+ Researchers

Spoke	Coordinator	Public Affiliates	Private Affiliates
1. Urban regeneration (City of tomorrow)	**UNIMIB**	POLIMI, UNIMI	UNIBOCCONI, ENI, EDISON, PIRELLI, THALES
2. Big Data-Open Data in Life Sciences	**UNIMI**	POLIMI, UNIMIB	UNIBOCCONI, ALMAVIVA, TIM, NOVARTIS, BIO4DREAMS, ASTRAZENECA, BRACCO
3. Deep Tech: Entrepreneurship & Technology Transfer	**POLIMI**	UNIMI, UNIMIB	UNIBOCCONI, FONDAZIONE POLIMI, OI, CAMOZZI, HUAWEI, INFINEON, A2A, HUMANITAS
4. Economic impact and sustainable finance	**UNIBOCCONI**	POLIMI, UNIMIB, UNIMI	FBK, RCS
5.Sustainable Fashion, Luxury and Design	**POLIMI**	UNIMI, UNIMIB	UNIBOCCONI, LUMSON
6. Innovation for Sustainable and Inclusive Societies	**UNIMIB**	POLIMI, UNIMI	UNIBOCCONI, EDISON, OI, THALES

PNRR MUSA

- **MUSA – Multilayered Urban Sustainability Action**
- PNRR Innovation Ecosystem
- 6 Spoke – 24 Partner - 970+ Researchers
- **Spoke 2 :** Technologies and processes for lifecycle management of life sciences and biomedical data

Spoke	Coordinator	Public Affiliates	Private Affiliates
1. Urban regeneration (City of tomorrow)	**UNIMIB**	POLIMI, UNIMI	UNIBOCCONI, ENI, EDISON, PIRELLI, THALES
2. Big Data-Open Data in Life Sciences	**UNIMI**	POLIMI, UNIMIB	UNIBOCCONI, ALMAVIVA, TIM, NOVARTIS, BIO4DREAMS, ASTRAZENECA, BRACCO
3. Deep Tech: Entrepreneurship & Technology Transfer	**POLIMI**	UNIMI, UNIMIB	UNIBOCCONI, FONDAZIONE POLIMI, OI, CAMOZZI, HUAWEI, INFINEON, A2A, HUMANITAS
4. Economic impact and sustainable finance	**UNIBOCCONI**	POLIMI, UNIMIB, UNIMI	FBK, RCS
5.Sustainable Fashion, Luxury and Design	**POLIMI**	UNIMI, UNIMIB	UNIBOCCONI, LUMSON
6. Innovation for Sustainable and Inclusive Societies	**UNIMIB**	POLIMI, UNIMI	UNIBOCCONI, EDISON, OI, THALES

Spoke 2 - Mission

- MUSA Spoke 2's focus is the design, implementation and deployment of a highly **innovative, secure ICT infrastructure and platform for Big Data collection and sharing**, suitable for both telemedicine and life science applications
- MUSA Spoke 2's objective is to **enable healthcare organizations to leverage secure data management capabilities and advanced AI-based analytics** to improve clinical practice, wellness and to deliver richer insights to internal and external data consumers

Prof. Ernesto Damiani
Dept. of Computer Science
Università degli Studi di Milano
co-PI of MUSA Spoke 2

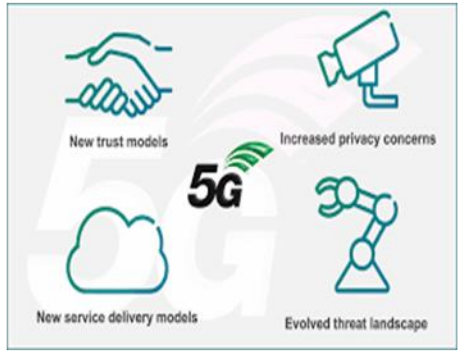
Prof. Gianvincenzo Zuccotti
Dept. of Biomedical and Clinical Sciences
Università degli Studi di Milano
co-PI of MUSA Spoke 2

Prof. Claudio A. Ardagna
Dept. of Computer Science
Università degli Studi di Milano
MUSA Spoke 2 – WP1 Leader

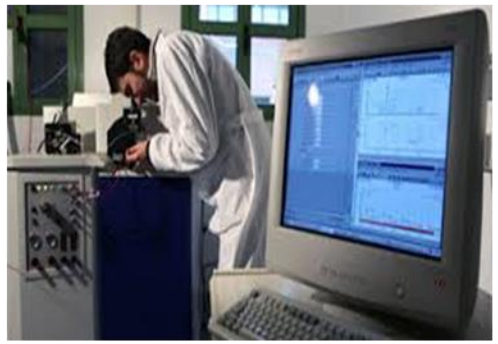
Prof. Marco Anisetti
Dept. of Computer Science
Università degli Studi di Milano
MUSA Spoke 2 – Task1.1 Leader

Dr. Nicola Bena
Dept. of Computer Science
Università degli Studi di Milano
MUSA Spoke 2 – Participant

MUSA Spoke 2 Four Pillars



Smart Devices (WP4)
Secure (mobile) communication (WP1)



Biomedical Data (WP1) Telemedicine services (WP2) Disease prevention (WP3)

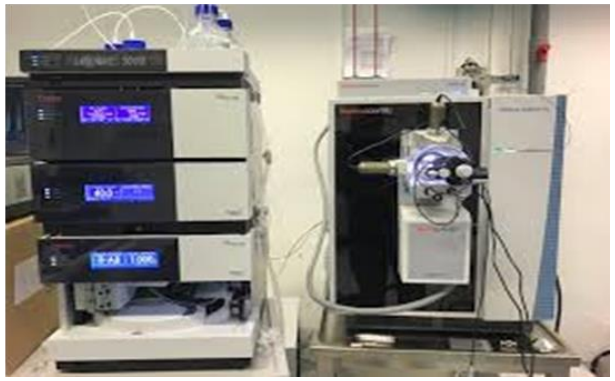
AI/ML services (WP1) Big Data Analytics (WP3)

Entrepreneurship and Innovation (WP5)
Business and Value Generation (WP1)

Spoke 2 – Outlook to the future

MIND

EVERYWHERE



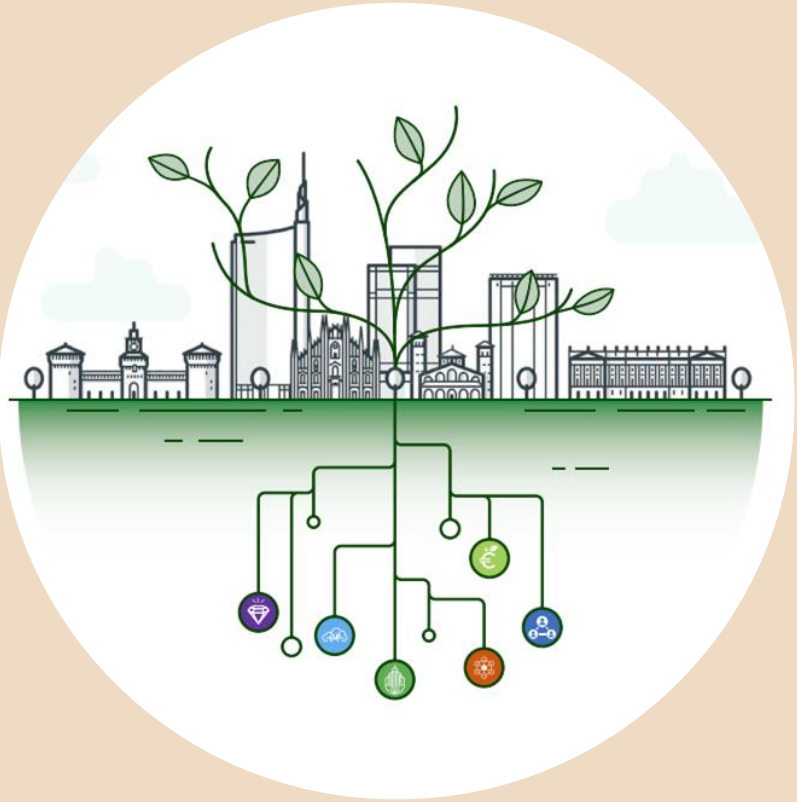
- **The 2023-2025 evolution of MUSA Spoke 2 will cover the entire data value chain**
- **We will support remote assistance, intelligent devices, 5G/6G communication, cybersecurity, artificial intelligence models, edge/cloud big data pipelines, human-system interfaces**

PNRR MUSA

MUSA digital platform for data analytics pipeline management in the cloud-edge continuum

Claudio Ardagna
Marco Anisetti

Università degli studi di Milano

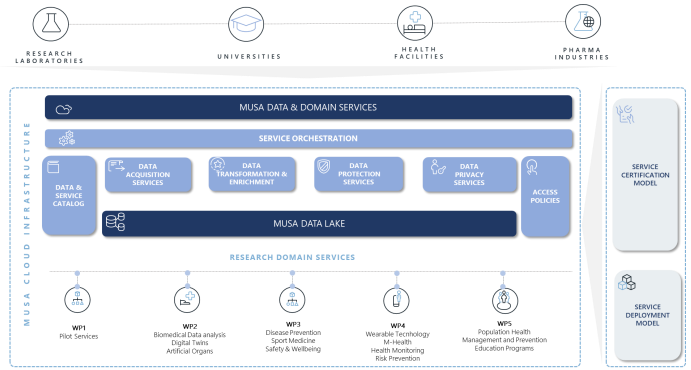


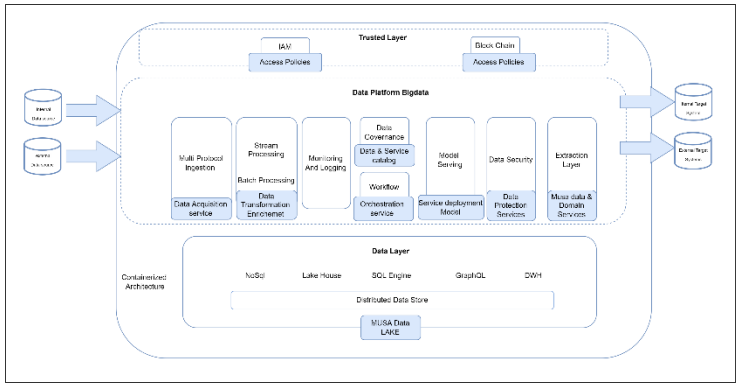
Objectives

- **Objective 1:** Development of an **innovative digital platform** for the rapid and secure storage and exchange of big data
  - Edge-cloud platform enabled by 5G
- **Objective 2:** Definition and design of an ecosystem of **services for data processing and analysis**
  - Support for the provision of advanced services to citizens
  - Provide value-added services that allow companies to build new services and business value
  - Based on AI techniques
  - Support secondary reuse of data for multi-stakeholder clinical studies led by pharmaceutical companies
- **Objective 3:** Definition of **advanced deployment solutions** for the provision of intelligent and accurate services with verified non-functional properties
- **Objective 4: Certified ecosystem that provides certified solutions and services**

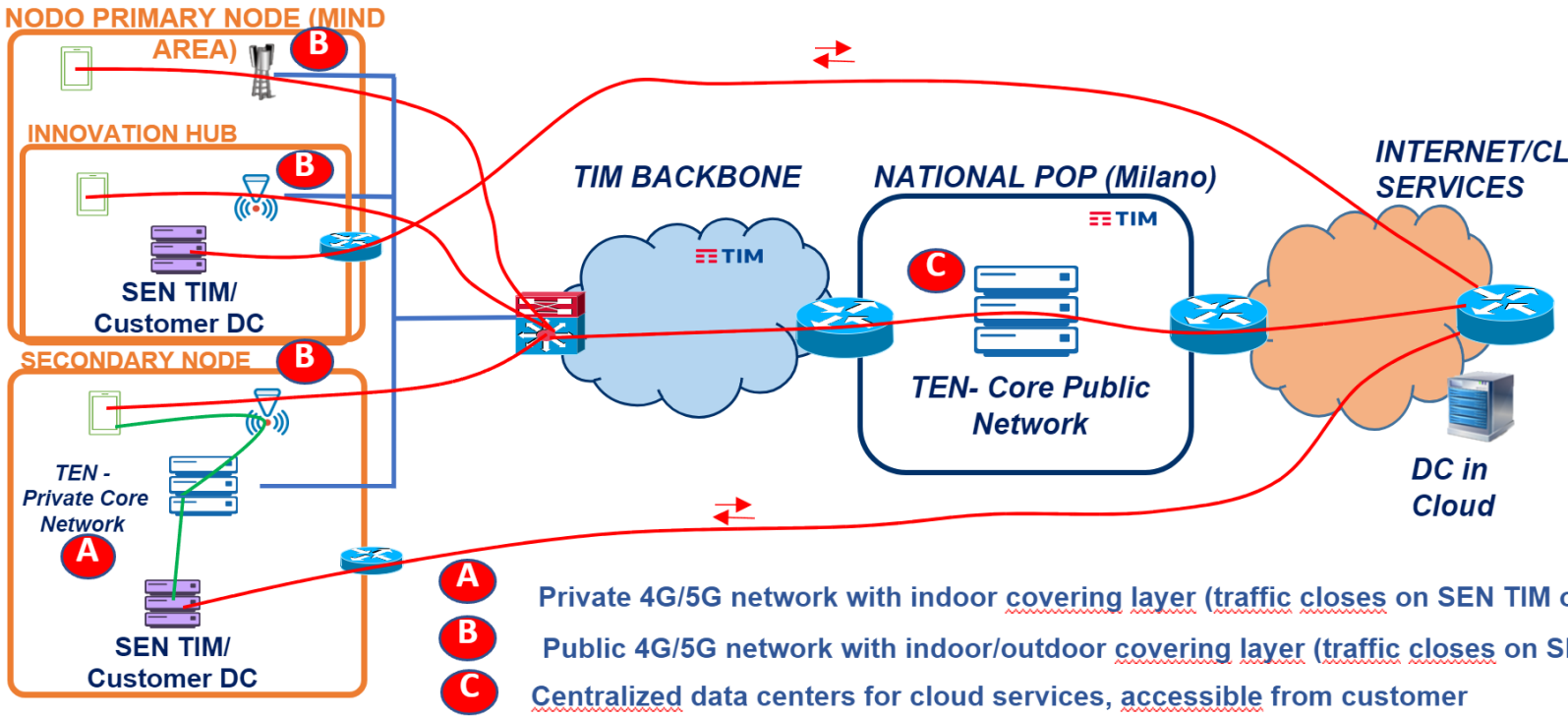
A holistic, innovative digital architecture for the storage and safe exchange of life sciences big data

- Design and deployment of an **innovative digital platform for data analysis and exchange** based on edge-cloud continuum and supported by a private 5G network and AI services
  - Data architecture and service selection
  - **MUSA Infrastructure development**
    - MUSA Cloud infrastructure build on kubernates (public cloud on AWS)
    - Complete simulator of a 5G infrastructure
  - Smart service deployment approach on a multi-platform environment driven by non-functional properties





Platform and Infrastructure: Future-proof, unique at Italian level, and one of the most advanced in Europe



Indoor 4G/5G NSA

Outdoor 4G/5G NSA

SEN: Edge Node Service
TEN: Telco Edge Node

Solutions At The Edge

5G on new Radio Frequences

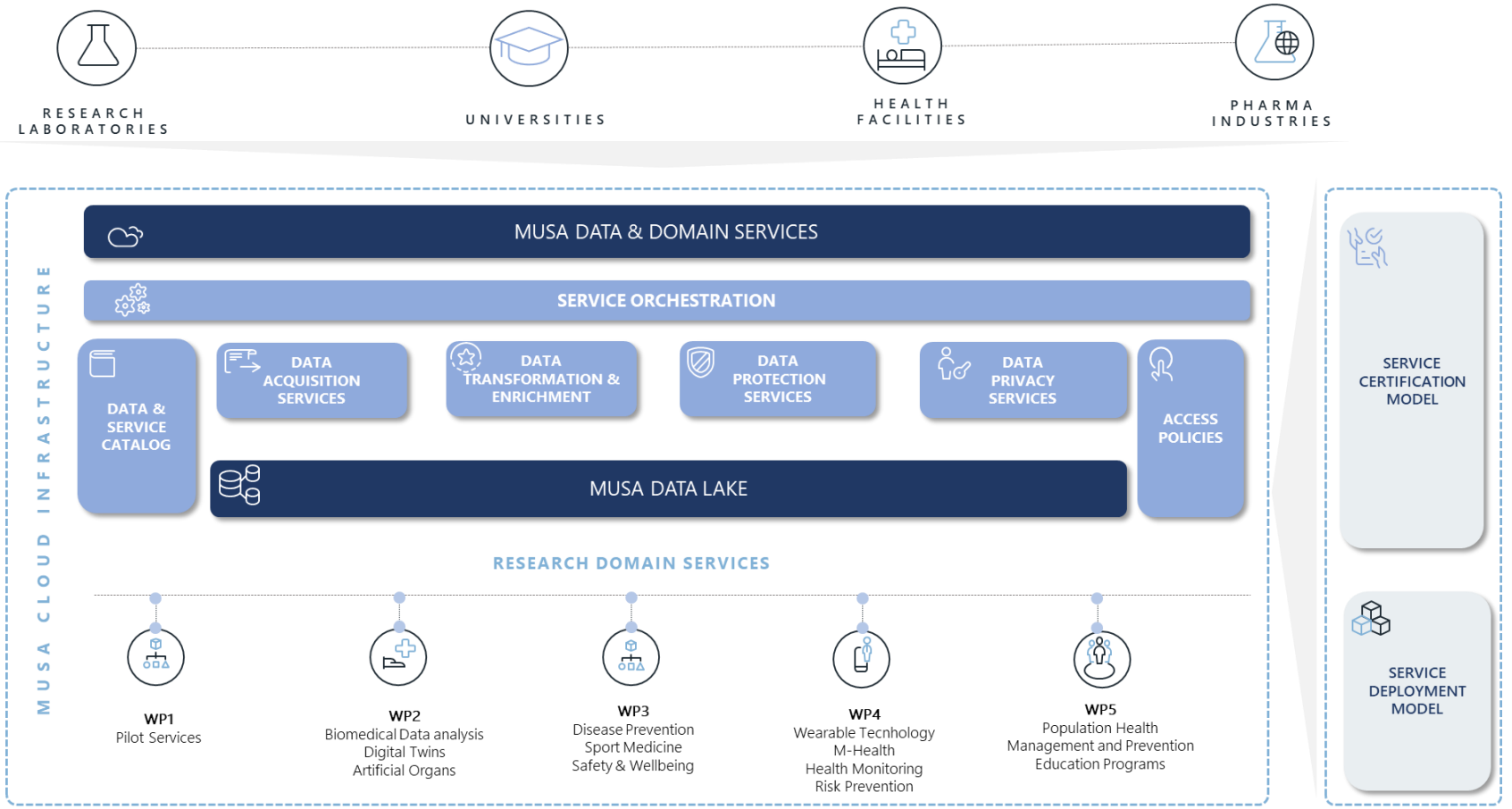
A Private 4G/5G network with indoor covering layer (traffic closes on SEN TIM or at customer-location DC)

B Public 4G/5G network with indoor/outdoor covering layer (traffic closes on SEN TIM or at customer-location DC)

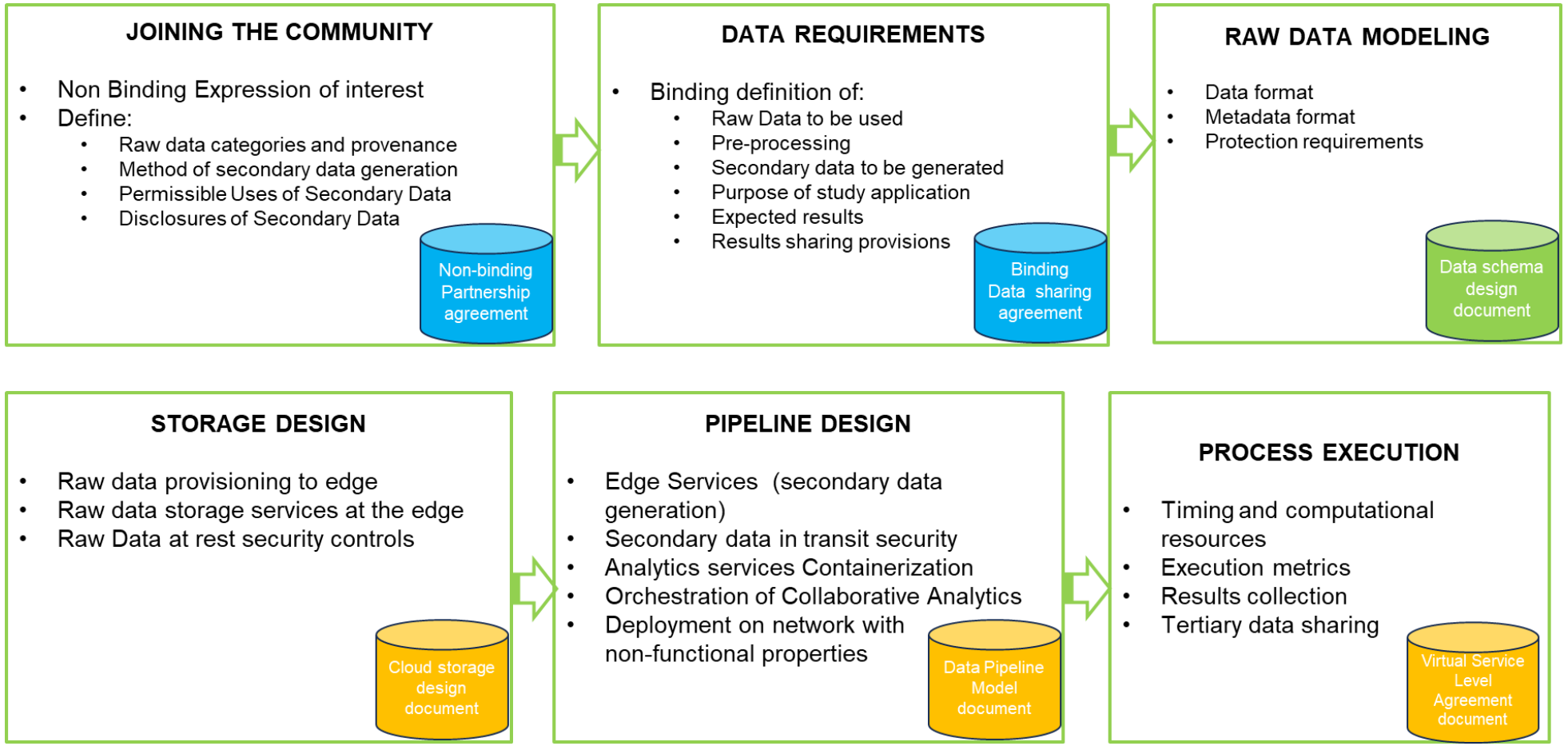
C Centralized data centers for cloud services, accessible from customer locations

Cloud Hub Data Architecture

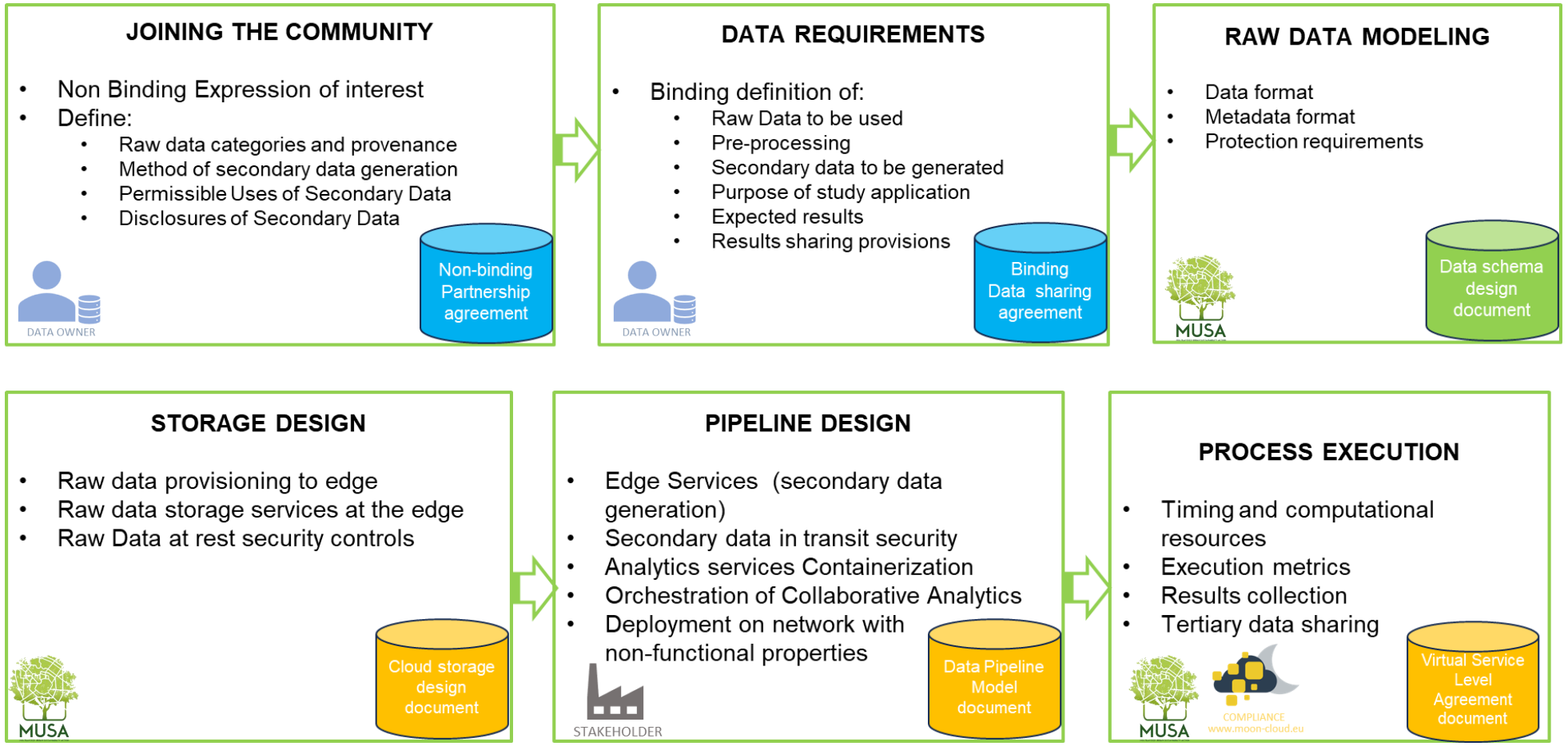
- Data lake
- Data Services
- Access Services



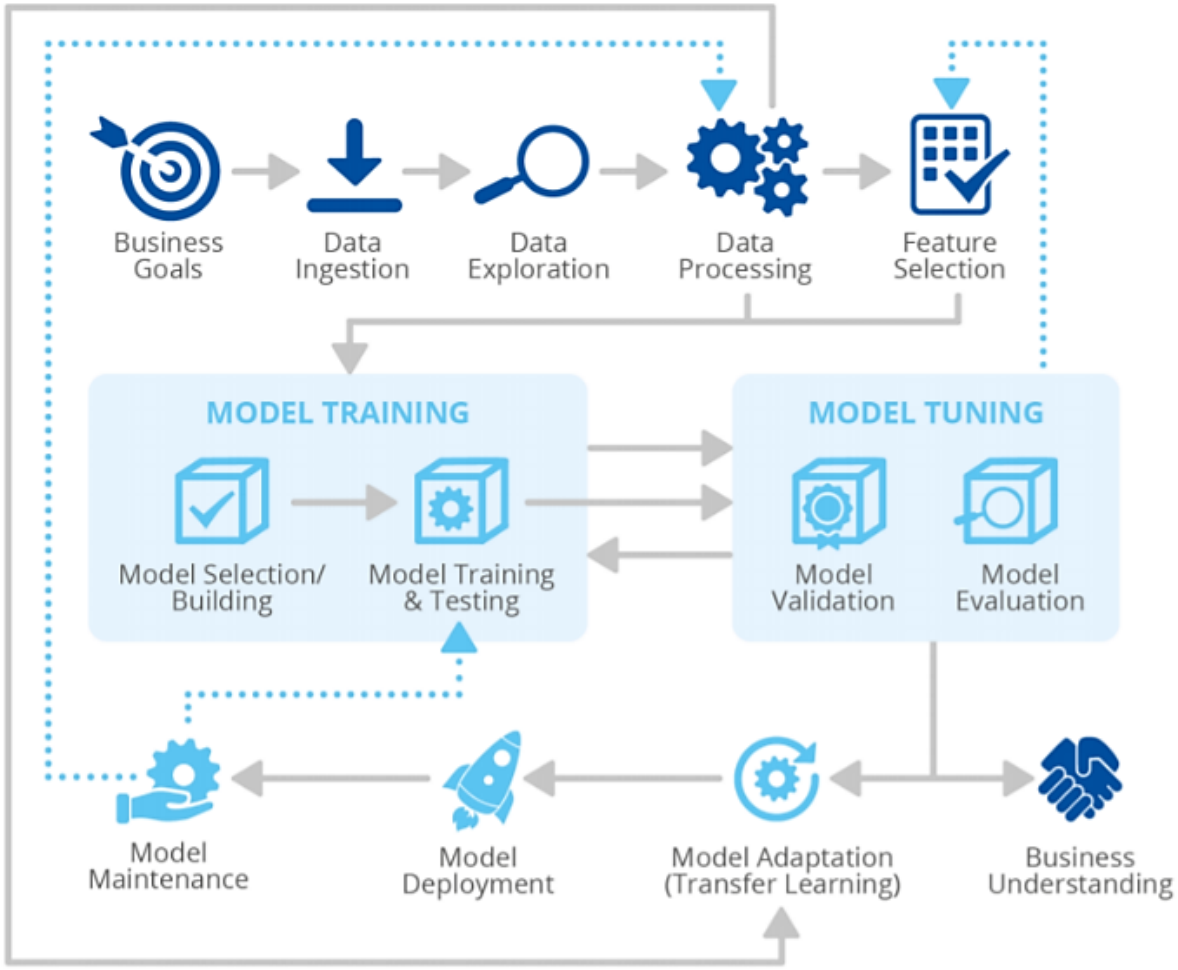
The MUSA Process



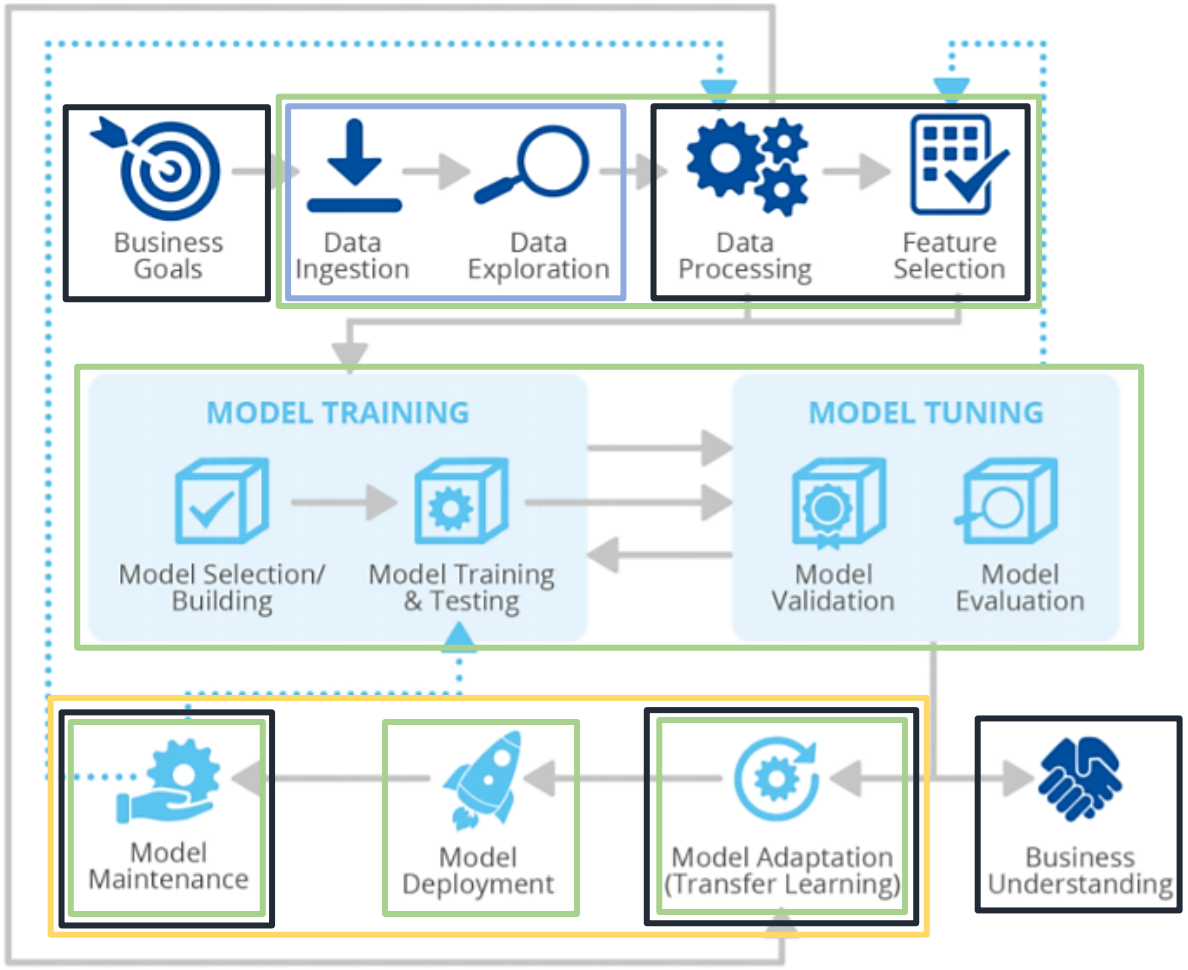
The MUSA Process



The complete life cycle of the data analytics platform



The complete life cycle of the data analytics platform



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
MNESSAA

Pilot Applications: Polycentric Studies

BRACCO
LIFE FROM INSIDE

- Image/tabular data fusion for federated learning of diagnostic models
- Goals
  - Metadata standardization and quality improvement
  - Fusion of image data and tabular metadata
  - Training of local models
  - Orchestrated federated learning for sharing local model parameters
  - Automatic feature selection and model comparison
- **Requirements:** data privacy
- **Modeling and deployment:** On-premises node with sensitive data; federated learning pipeline with centralized cloud node

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Pilot Applications: Wellbeing and disease prevention

- Quantitative validation of good practices
- Goals
  - Definition of correct lifestyle (metrics and impact indices)
  - Validate whether a choice has a positive or negative impact on wellness
  - Validation at different levels: urban planning, nutrition…
  - **Requirements:** privacy of input data; quality of data processing
  - **Modeling and deployment:** on-premise node for the preparation of sensitive data, cloud pipeline with direct ingestion of processed data

MUSA

Pilot Applications: Diagnostic and therapeutic services and risk models

- Diagnostic services and specialist tests with remote doctor or mobile patient
  - Diagnostic tests in the ambulance
  - Ultrasound scans with a medical examiner not present on site (for mountain communities)
  - **Requirements:** low latency, privacy and security of communication channels
  - **Modeling:** pipeline with 5G Edge nodes
- Management of massive IoT scenarios
  - Risk models for widespread diseases (cardiac risk)
  - Optimization of processes and consumption in smart cities (mobility)
- Data protection in distributed, heterogeneous and mobile environments
  - Throughout the data lifecycle

MUSA

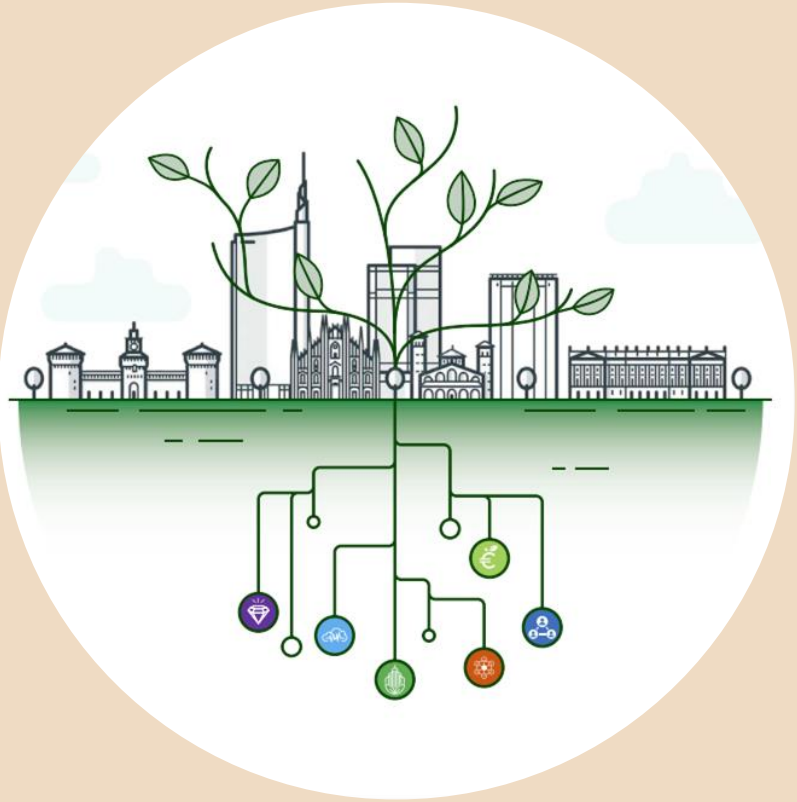
www.moon-cloud.eu

PNRR MUSA

Smart Service Deployment Solution in Edge-Cloud Continuum

Claudio Ardagna
Marco Anisetti

Università degli studi di Milano



MUSA Edge Cloud Continuum

- Edge-Cloud Continuum: a novel paradigm enabling new opportunities for distributed applications
- **MUSA Continuum involves 5G MEC**
  - fully support compositions of services
  - suit advanced QoS requirements
    - vast computational resources: Cloud
    - low-latency: MEC or far Edge
    - legacy systems compatibility/strict privacy: Edge on-Premises
  - differentiate security/privacy policies

Challenges

- C1 - Compose and distribute both MUSA' services and customers' services
- C2 - Maintaining advanced QoS such as security and privacy
- C3 - Involve facilities of different providers
- C4 - Coexistence with other services and compositions

Solutions: MUSA approach

- Automatically decide how to **compose** (C1) and where to **deploy** services (C3) guided by **QoS** requirements (C2)
- Cross-provider Continuum using **MUSA continuum agents** (C3)
- **Containerization** of services and assurance-based **monitoring** (C4)

Deployment: open issues

- Obtain scalable deployment in continuum
- **Guarantee properties** on the composed deployment
  - Some solutions have been proposed in serverless computation
    - limited to resource allocation (i.e., CPU, memory, and bandwidth)
    - cannot be applied to compositions
- MUSA deployment Goal: automatic QoS-aware deployment solution for composed services in the Continuum
- **Deployment decisions impact on QoS (e.g., latency)**

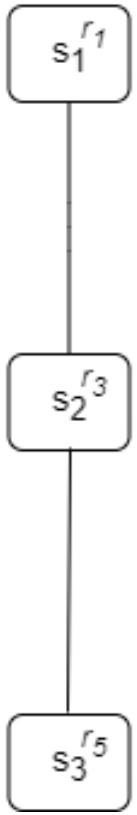
MUSA Scenario

- Different entities are part of the MUSA Continuum
  - CSP offering services and hosting facilities (mainly Cloud)
  - Telco providers offering Edge telco nodes and connectivity
  - MUSA clients willing to use MUSA
    - allowing their premises to join MUSA continuum
    - upload their services on the MUSA catalogue (private or public accessible)
- **Data** hosted on privacy-preserving MUSA Data lake or on MUSA-enabled on-premises data warehouse

MUSA Scenario

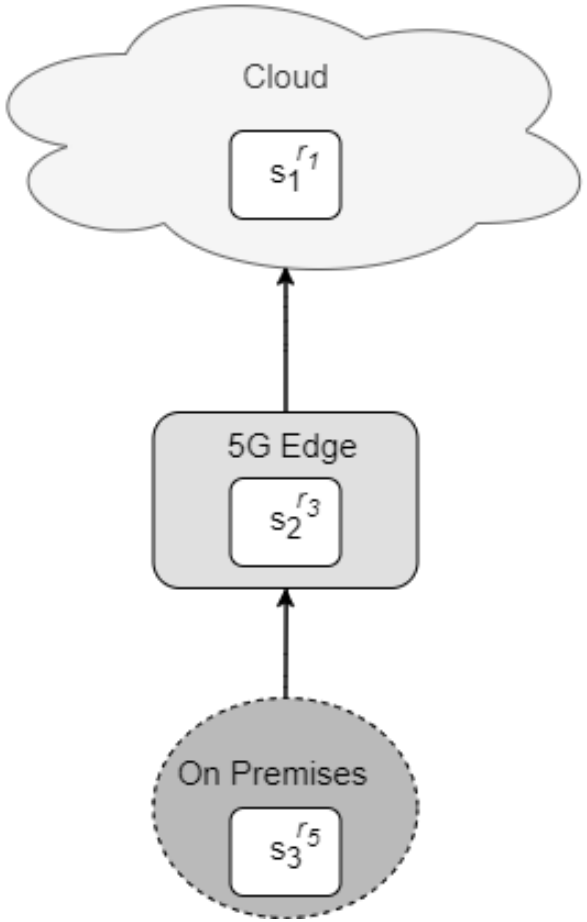
- Clients: express their workflow of services composing custom and/or MUSA services
  - Musa service catalogue contains general purposes MUSA services such as authentication, data preparation, basic AI algorithms…
- CSP and Telco: have an agreement with MUSA to offer their landing facilities for the client's workflows to be deployed and executed

MUSA Scenario



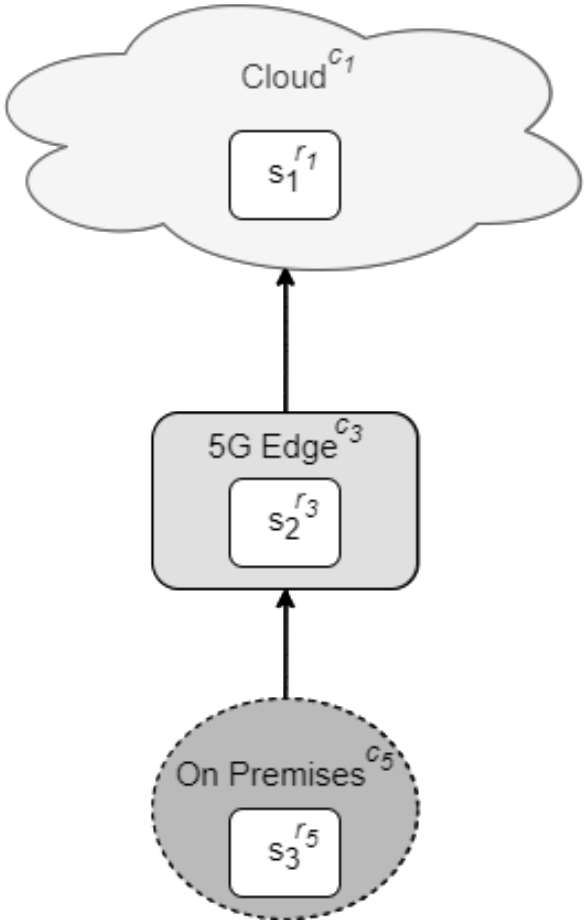
- Clients: **design** workflows of services s_i with QoS requirements r_i

MUSA Scenario



- Clients: design workflows of services s_i with QoS requirements r_i
- CSPs: offer Cloud deployment facilities
- Telco operators: offer 5G Edge and core network capabilities
- Clients: allows on-premises deployment

MUSA Scenario



- Clients: design workflows of services s_i with QoS requirements r_i
- CSPs: offer Cloud deployment facilities
- Telco operators: offer 5G Edge and core network capabilities
- Clients: allows on-premises deployment
- Facilities: provide specific **capabilities** c_i

Deployment Requirements

- **R_1- continuum-readiness**: seamlessly deploy services on every Continuum premises
- **R_2 - property-driven**: grounded on QoS properties and *constraints* expressed by the client
- **R_3 - technology agnostic**: handle heterogeneous deployment facilities
- **R_4 - comprehensive model**: provide a general way to represent workflows and facilities
- **R_5 - interoperability**: able to interact with different CSP/Telco hooks
- **R_6 - context adaptability**: automatically manage deployment life-cycle

Deployment Requirements: State of the Art

Most of the solutions are focused on resources (CPU, storage, Memory)

Difficulties in addressing Requirements

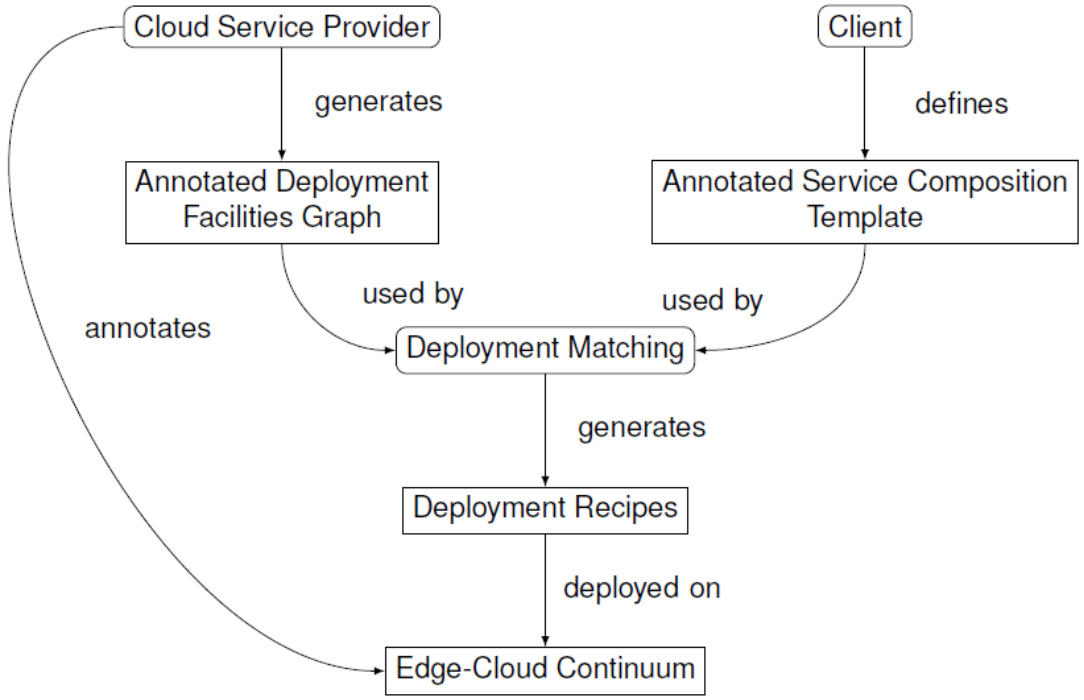
- **R_2 - property-driven**
- **R_6 - context adaptability**

ML solutions increasingly adopted

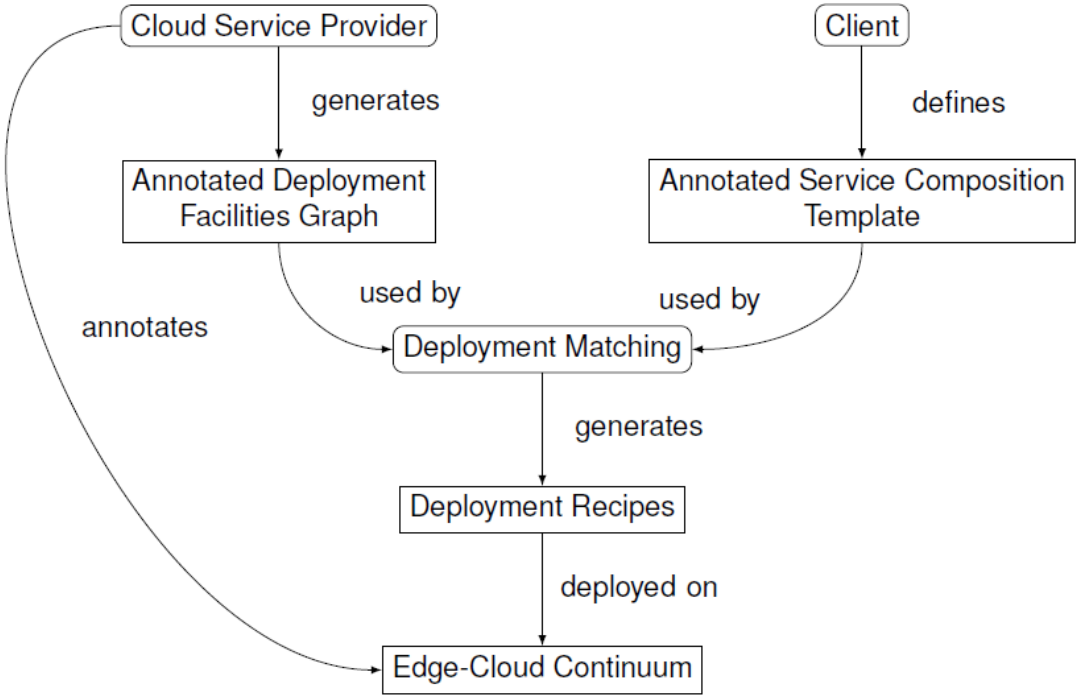
Author	Ref.	R1	R2	R3	R4	R5	R6
K. Fu et al.	[12]	✓			✓		~
A. Orive et al.	[13]	✓	~		✓	✓	
A. Brogi et al.	[15]		~	✓	✓		
V. Casola et al.	[14]	✓	✓	✓		✓	
S. Nastic et al.	[9]	✓		✓			
N. Akhtar et al.	[6]		~	✓		✓	~
A. Das et al.	[8]			✓	✓	✓	
M. Anisetti et al.	[16]		✓	✓			
J. Quenum et al.	[10]		✓	✓		✓	
Our Work		✓	✓	✓	✓	✓	✓

M. Anisetti, F. Berto, R. Bondaruc, ``QoS-aware Deployment of Service Compositions in 5G-empowered Edge-Cloud Continuum,'' IEEE CLOUD 2023

Methodology

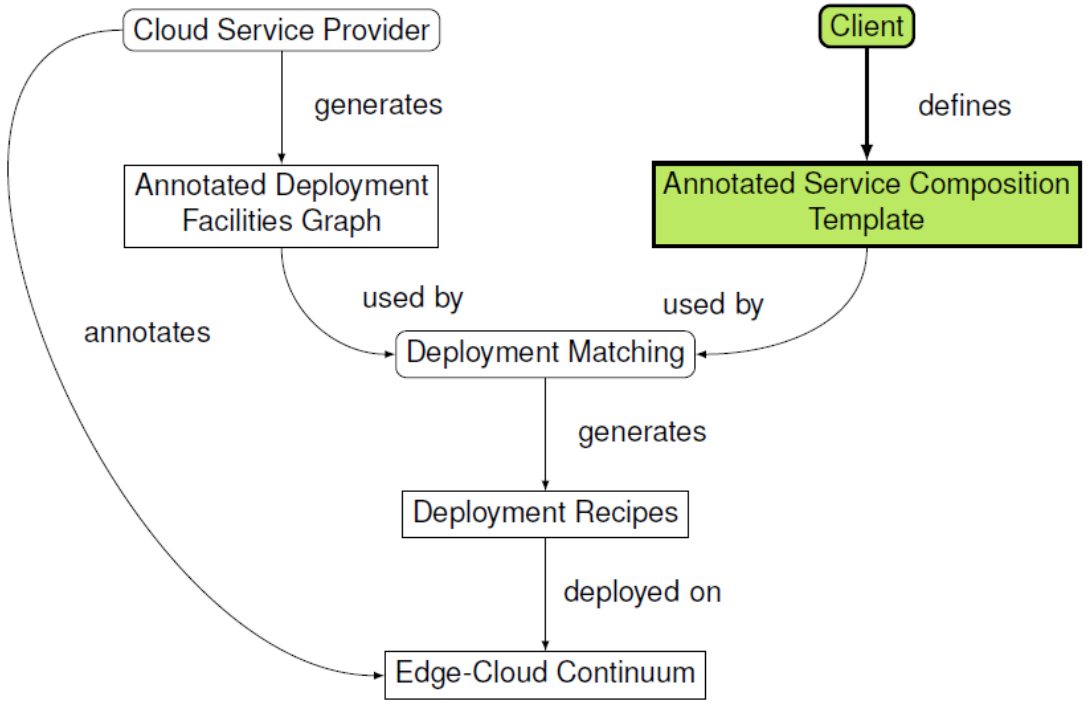


Methodology



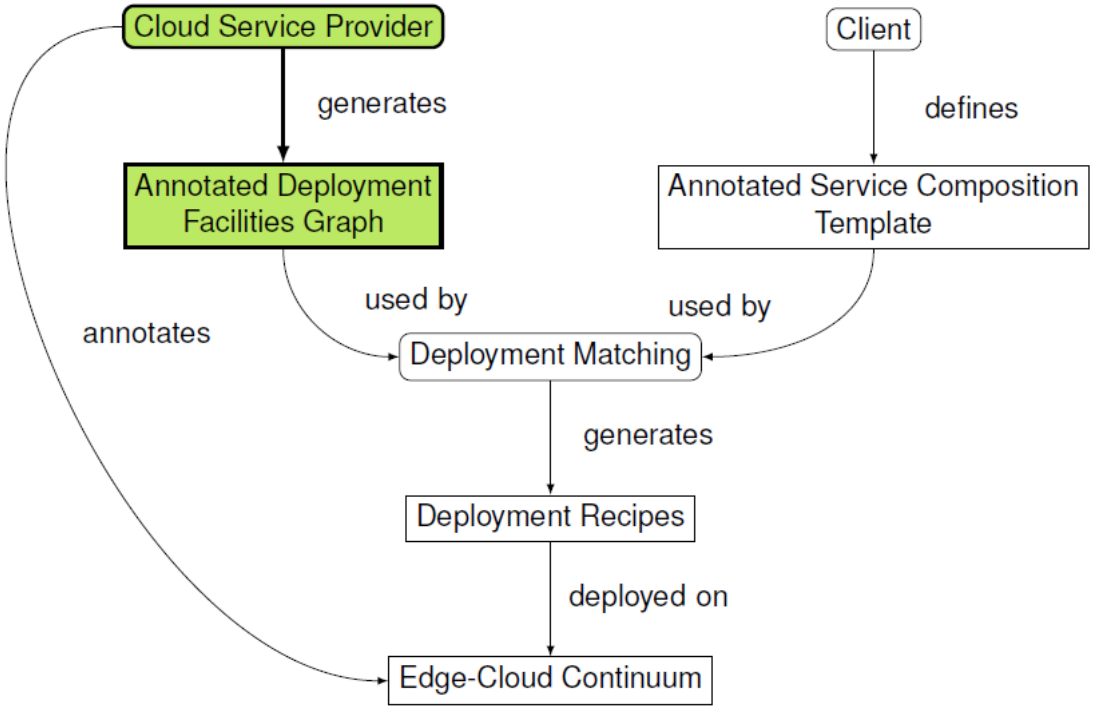
- Starting from annotated models

Methodology



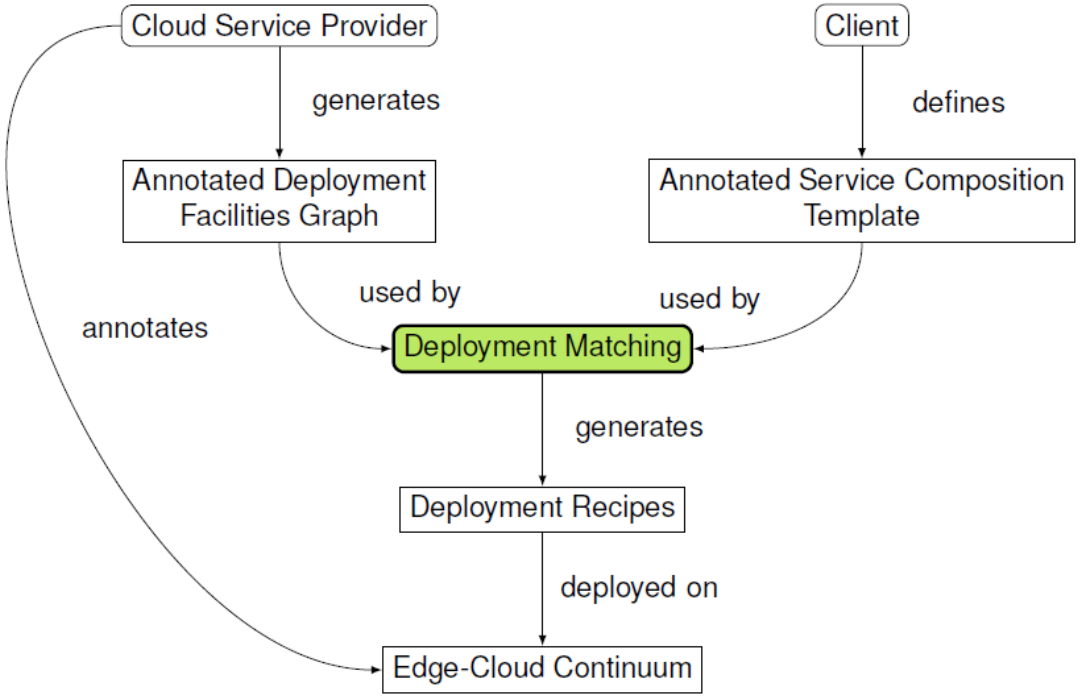
- Starting from annotated models
  - service composition template (client)

Methodology



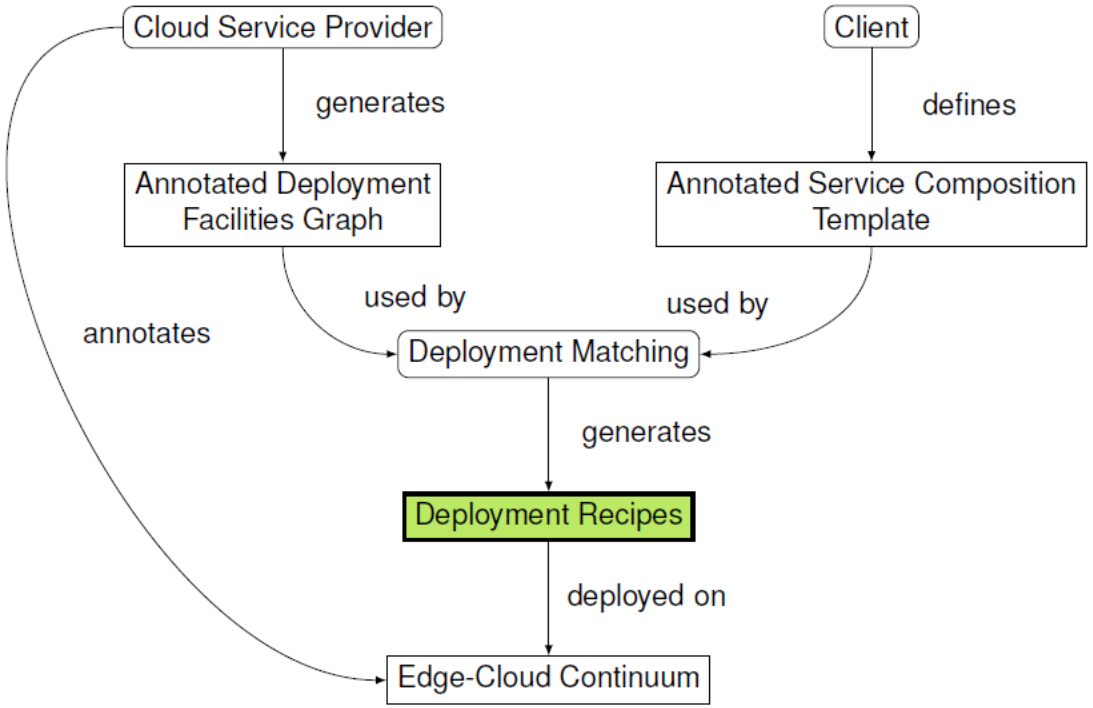
- Starting from annotated models
  - service composition template (client)
  - deployment facilities graph (CSP)

Methodology



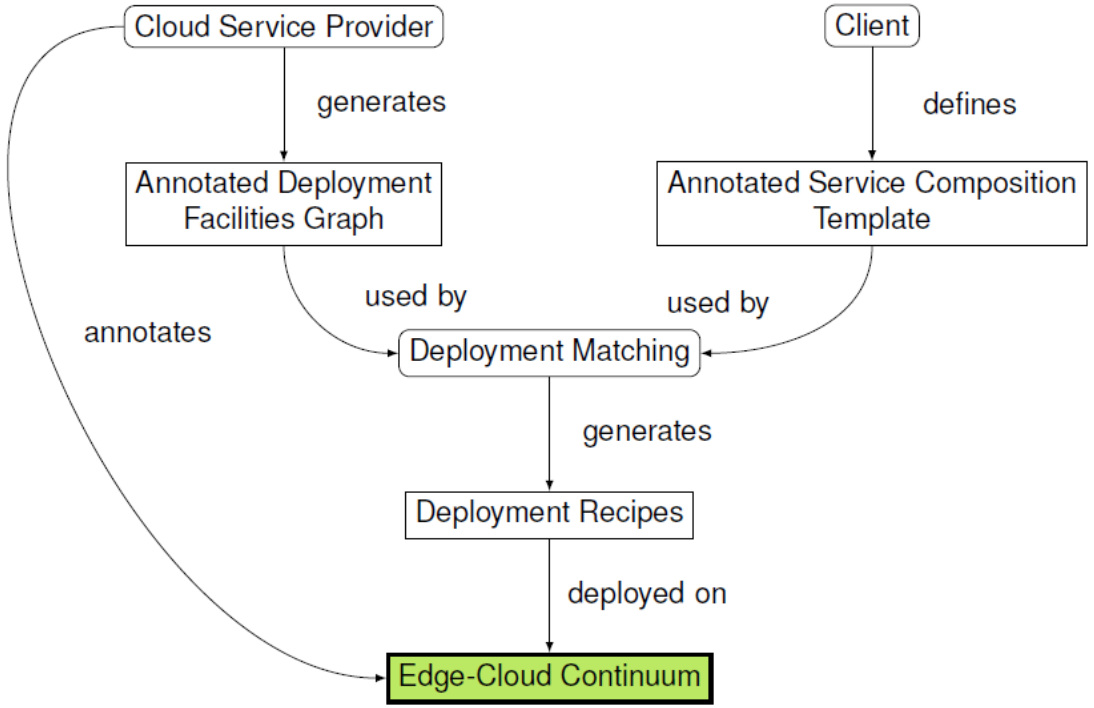
- Starting from annotated models
  - service composition template (client)
  - deployment facilities graph (CSP)
- Match between requirements and constraints (services) and capabilities (facilities)

Methodology



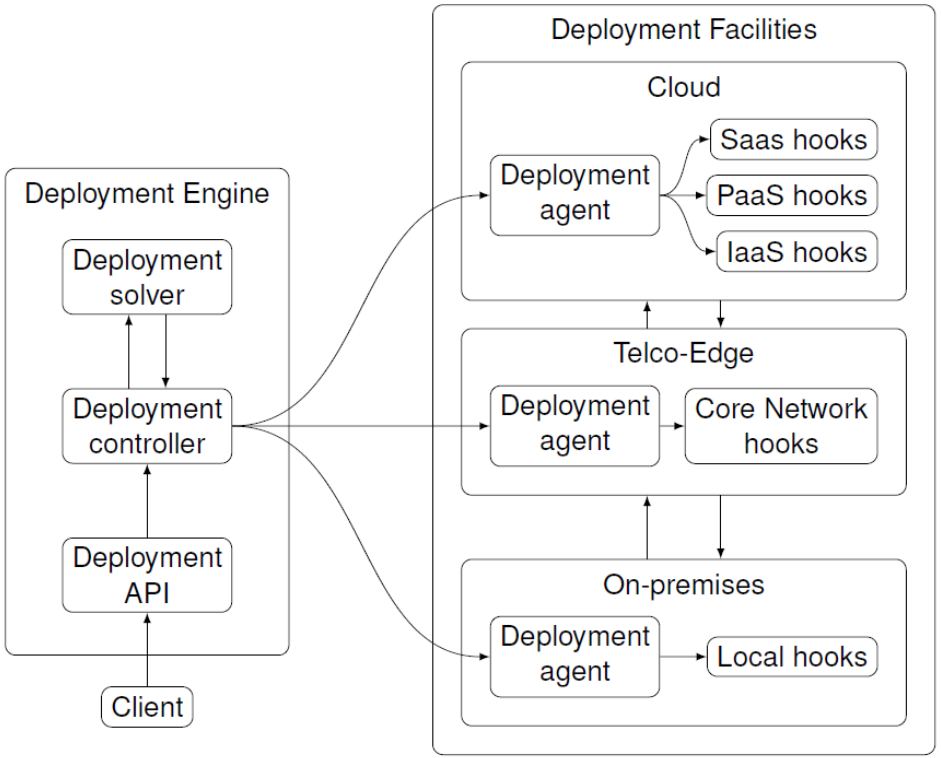
- Starting from annotated models
  - service composition template (client)
  - deployment facilities graph (CSP)
- Match between requirements and constraints (services) and capabilities (facilities)
- Generate a deployment recipe

Methodology

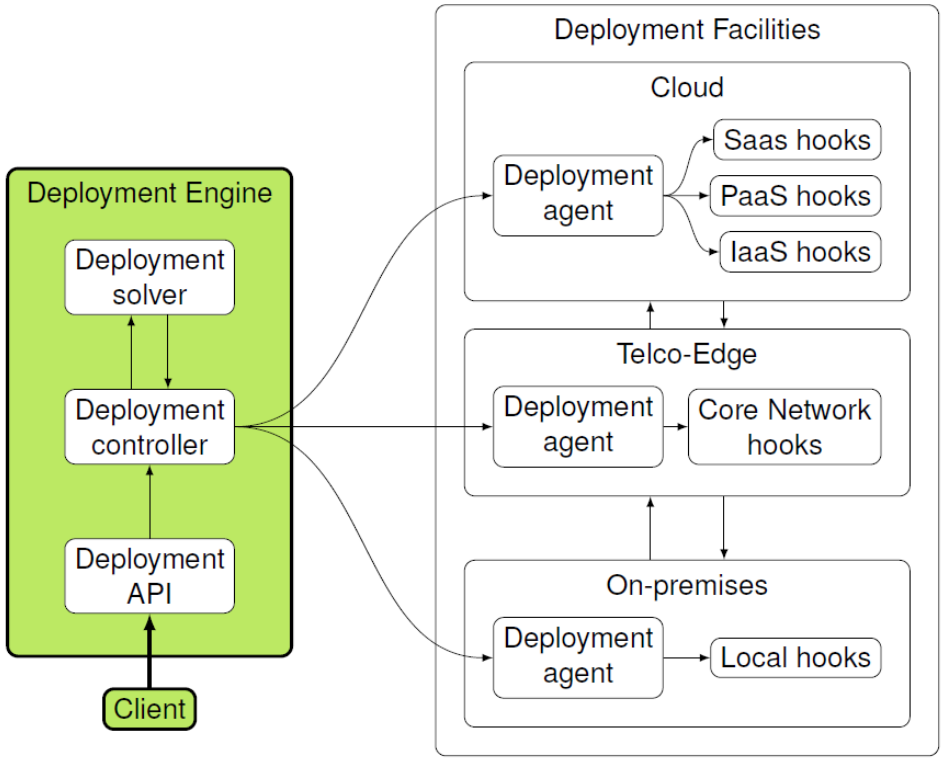


- Starting from annotated models
  - service composition template (client)
  - deployment facilities graph (CSP)
- Match between requirements and constraints (services) and capabilities (facilities)
- Generate a deployment recipe
- Concretize recipe for each specific node

Architecture

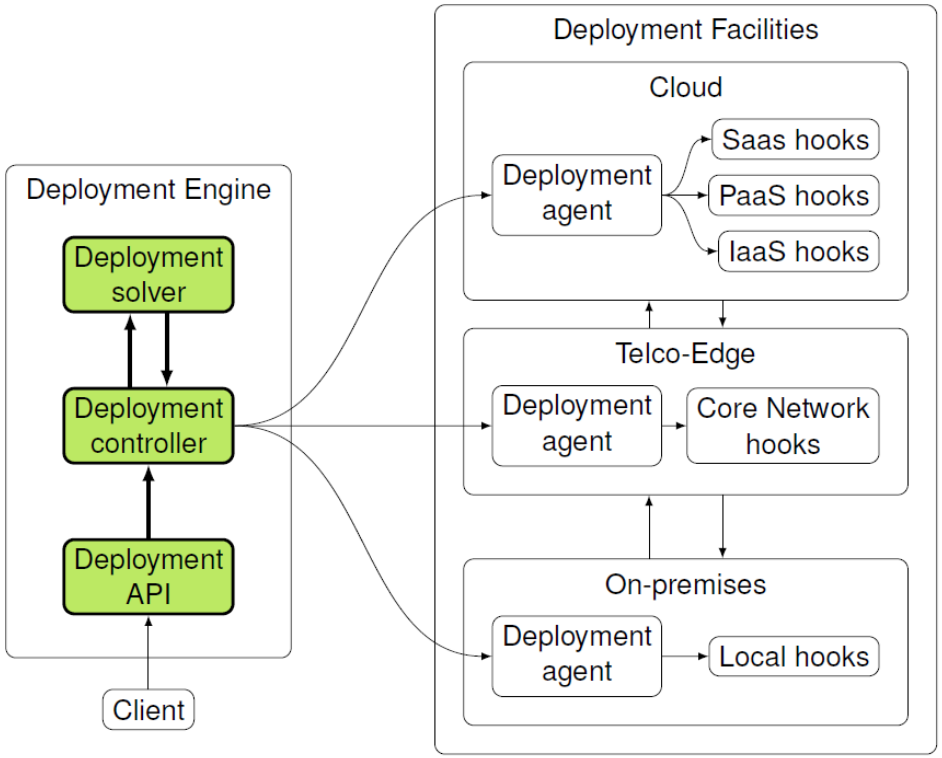


Architecture



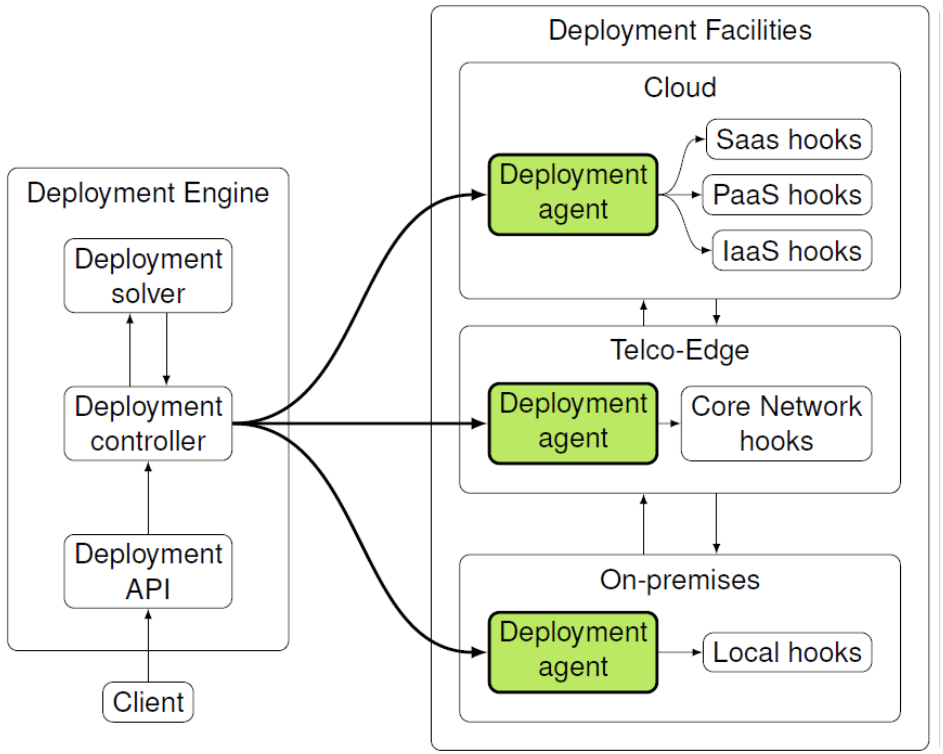
- Client requests deployments via Deployment Engine

Architecture



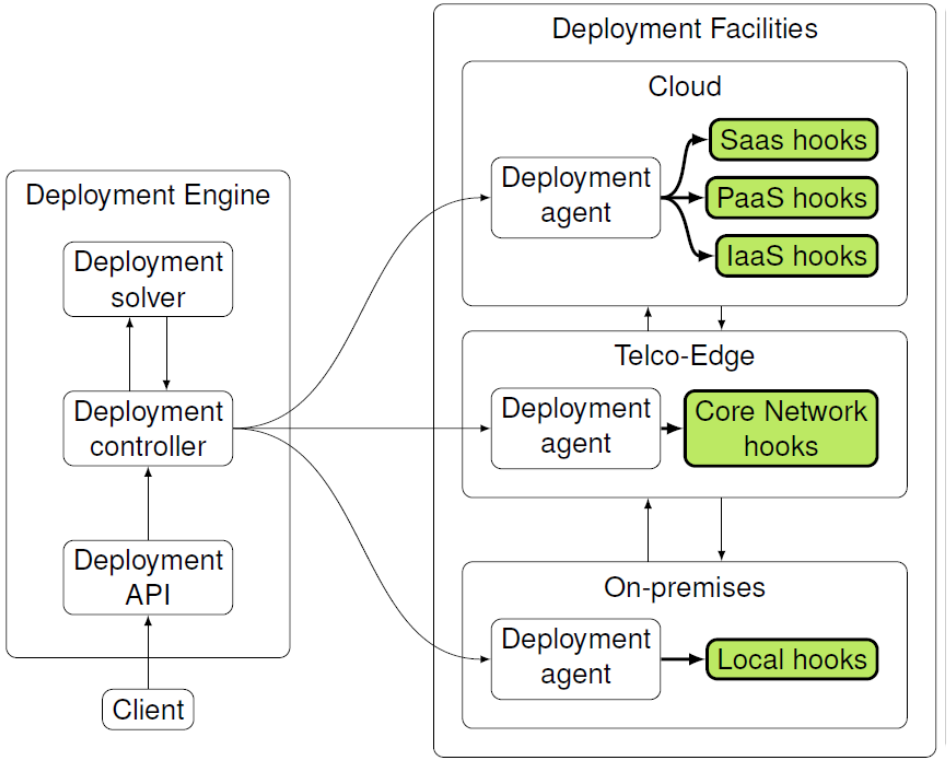
- Client requests deployments via Deployment Engine
  - matches services with facilities through the Deployment solver

Architecture



- Client requests deployments via Deployment Engine
  - matches services with facilities through the Deployment solver
  - targets facilities through Deployment Agents

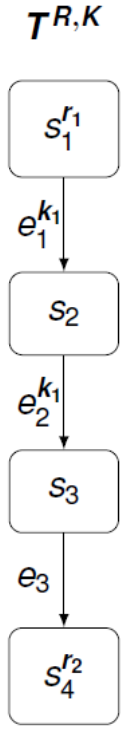
Architecture



- Client requests deployments via Deployment Engine
  - matches services with facilities through the Deployment solver
  - targets facilities through Deployment Agents
- Facilities provide hooks to relevant resources or services

Model of workflows and facilities

- **Annotated Service Composition Template** $T^{R,K}$ is
  - a directed graph T = (S, E)
  - annotated with requirements r and constraints k



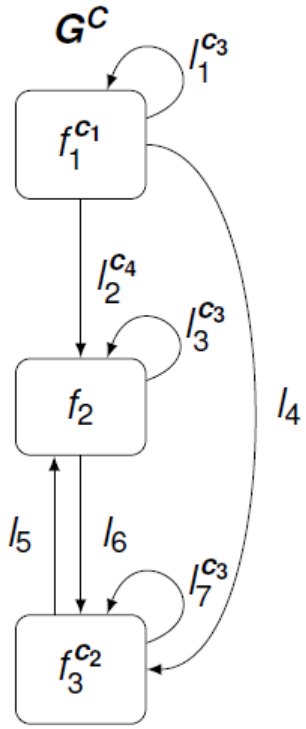
r_1 = (*Confidentiality, Isolation*)

r_2 = (Integrity, Rest)

k_1 = (*Bandwidth, 200, ≥*)

Model of workflows and facilities

- **Annotated Deployment Facilities Graph** G^c is
  - a directed graph $G = (F, L)$
  - annotated with capabilities c
  - implement capabilities



c_1 = (Confidentiality, Isolation, =, [])

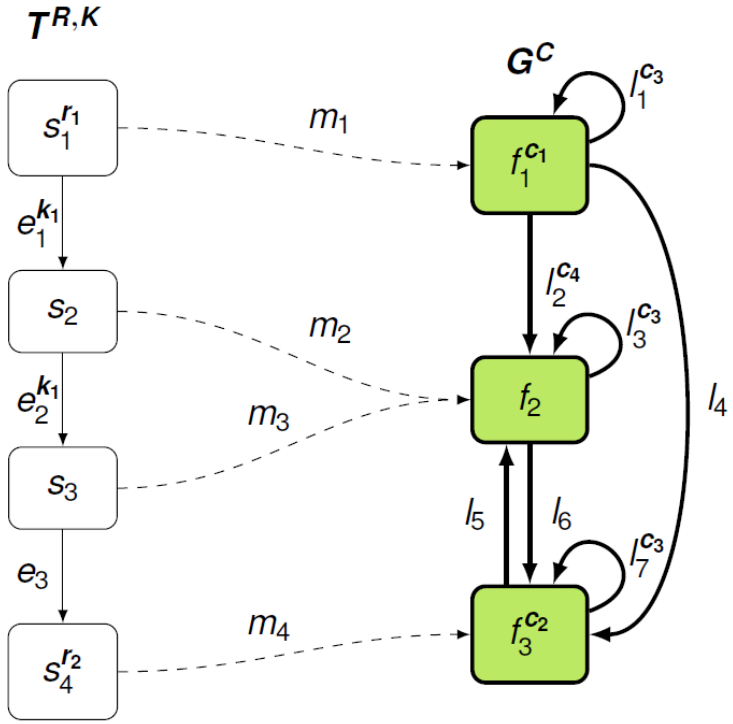
c_2 = (Integrity, Rest, =, [service: sI; mode: interception])

c_3 = (Bandwidth, +∞, =, [])

c_4 = (Bandwidth, 500, ≥, [])

Deployment Matching

- Searches for the most suitable solution for the QoS-aware deployment
  - takes as input the annotated models
  - generates a set of suitable solutions M
  - among them selects the one better satisfying a provider-specific policy
    - lowest operational cost $\{m_1,m_2,m_3,m_4\}$
- if M is empty, the deployment cannot take place under the specified requirements

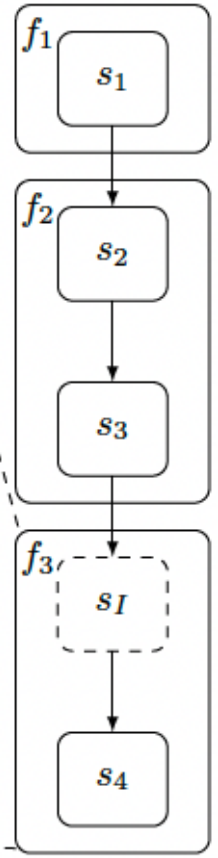


Deployment Recipes

Generates recipes to enable properties

- configure facilities
- deploy additional MUSA services to support properties
- deployment recipes **enriched** with hooks to enable properties
- consists of three parts
  - service deployment configuration
  - support MUSA services description
  - modality of integration (i.e., none, interception or wrapper)

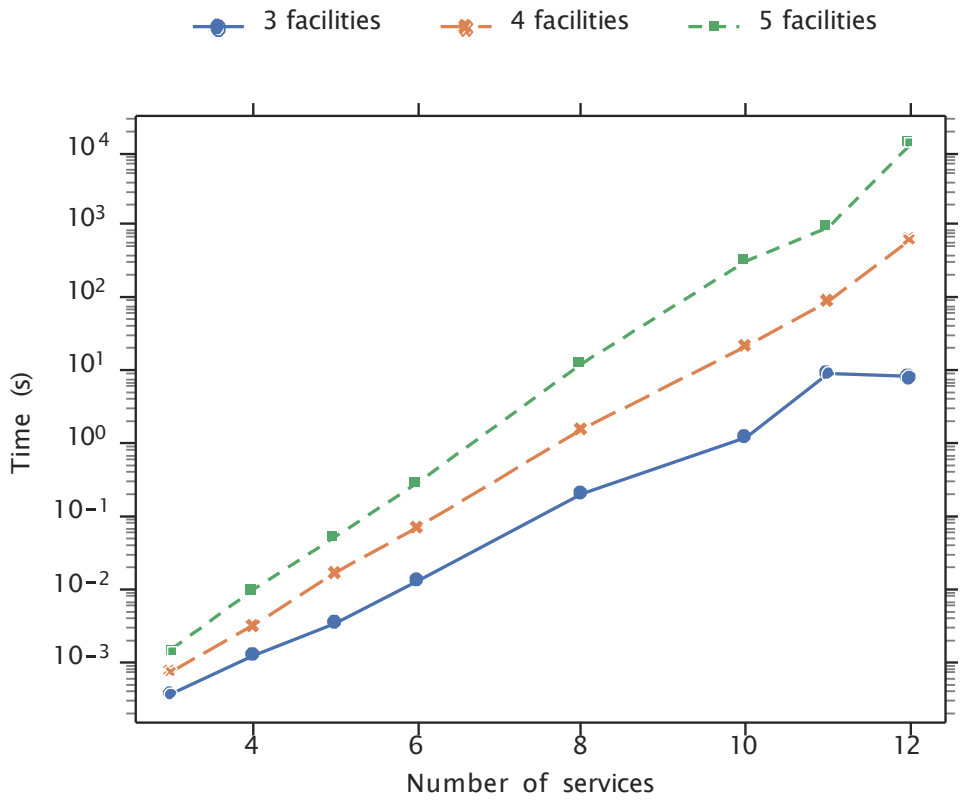
```
facility:
  name: f3
  services:
  - name: s4
    kubernetes_template:
      apiVersion: apps/v1
      kind: Deployment
      metadata:
        name: s4
      spec:
        replicas: 1
        template:
          spec:
            containers:
            - name: s4
              image: local.registry/s4
              ports:
              - containerPort: 80
    capabilities:
    - property: integrity
      attribute: rest
      operation: "="
      implementation:
        service: sI
        mode: interception
```



Matching Evaluation

The matching was evaluated in a simulated MUSA continuum environment

- matching time of different sets of services and facilities
- matching performed in an exhaustive manner (no optimization)
- the number of services dominates the deployment time (exponential)
  - impractical after s = 10 and f = 5



Deployment Evaluation: MUSA Pilots

Different type of pilots to be deployed on the MUSA Continuum

- **Cloud only**: high performance processing pipelines
- **On-premises and Cloud**: privacy y preserving polycentric studies (e.g., using Federated Learning)
- **On-premises, 5G Edge and Cloud**: polycentric 5G connected studies, low latency consultancy services

Monitoring of the MUSA Cloud Facilities via moon cloud probes

Example of 5G MUSA pipeline

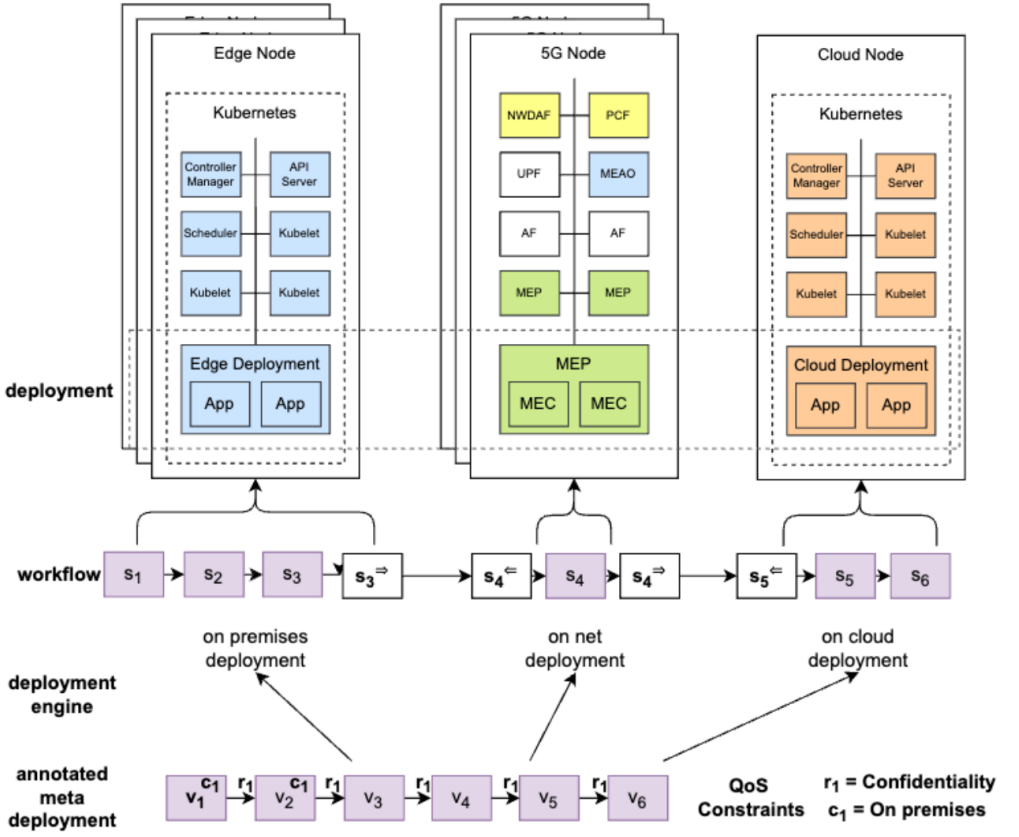
Custom and MUSA native services

Connectivity services inserted (e.g., s3⇒) by MUSA deployer

Data pre-processing on premises Edge

Low latency processing in 5G Edge

Computational-expansive analytics in Cloud

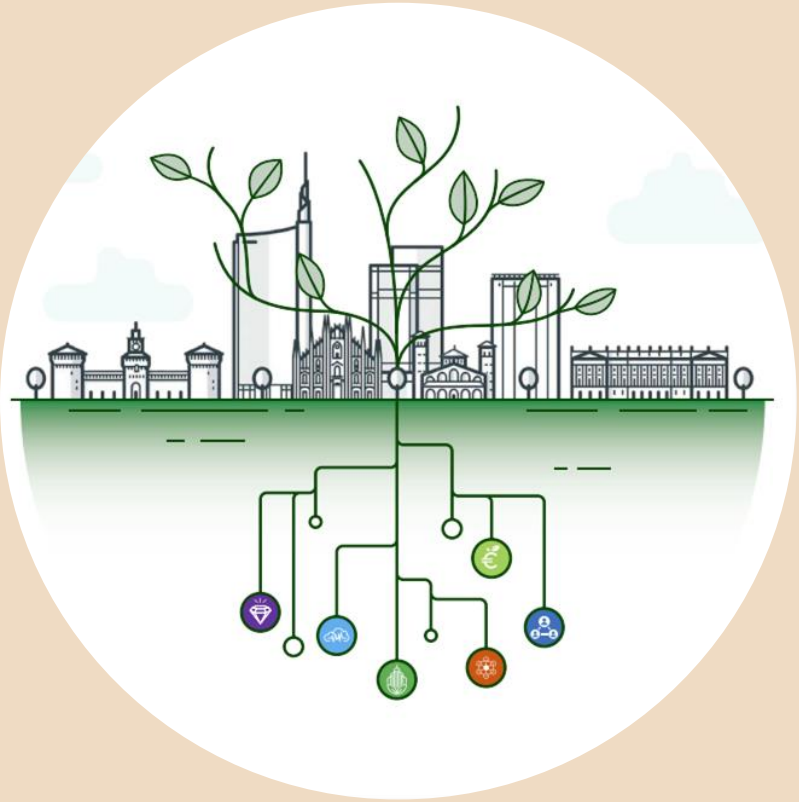


M. Anisetti, F. Berto, M. Banzi, ``Orchestration of data-intensive pipeline in 5G-enabled Edge Continuum,'' IEEE EDGE 2022

PNRR MUSA

What's next? Cascading grants

27 November 2023



Next Steps

1. Encapsulation and Deployment of Data-Intensive Services in Polymorphic Containers
2. Automatic optimization of the deployment and operation of services on the platform through Artificial Intelligence techniques
3. Preparation and management of an environment for testing and verification of services on the 5G infrastructure
4. Non-functional driven deployment of services
5. Certification of the data services

Next Steps

CASCADING GRANTS

1. Encapsulation and Deployment of Data-Intensive Services in Polymorphic Containers
2. Automatic optimization of the deployment and operation of services on the platform through Artificial Intelligence techniques
3. Preparation and management of an environment for testing and verification of services on the 5G infrastructure
4. Non-functional driven deployment of services
5. Certification of the data services

Encapsulation and Deployment of Data-Intensive Services in Polymorphic Containers

- Development and implementation of **new techniques for the rapid creation and implementation of polymorphic implementations of software services** starting from user code, through multiple multi-format containers.
- Development and implementation of technologies for **manual deployment, testing and compliance verification of polymorphic implementations of digital services** on cloud and mobile network edge.
- **User-friendly interface** for controlling and executing service implementation and configuration tasks

Automatic optimization of the deployment and operation of services on the platform through Artificial Intelligence techniques

- Development, training and testing on benchmark data of high-dimensional supervised computational learning models for the **automatic real-time distribution of software services on the 5G network**, capable of minimizing complex cost functions in terms of energy and latency.
- Adoption of **assurance frameworks** capable of applying policies and verifying the configuration of the necessary network resources, as well as the feasibility and non-functional properties of services and processes.

Preparation and management of an environment for testing and verification of services on the 5G infrastructure

- Creation of an **innovative environment for the testing and continuous monitoring of digital processes in the healthcare sector**, which also includes peripheral devices and edge computing services, in the form of a private mobile network composed of territorially distributed nodes
- Implementation of the **private multi-node mobile network** that hosts the test and monitoring environment on the fifth generation mobile network (5G) in the **Milan metropolitan area**
  - **MUST be in operation until the end of the MUSA project**
- Inclusion in the testing and monitoring environment of **at least two** geographical areas of interest for innovation in the biomedical field **in the Milan metropolitan area** with the possibility of simultaneous access of **at least five devices per area**
- Preparation and interface for the future extension of the test environment in a large-scale production environment in a metropolitan network
- Complementarity with existing mobile network test infrastructures

Cascading grant

- Published November 22, 2023

- https://work.unimi.it/servizi_ricerca/bandi_finanz/130446.htm