Finanziato dall'Unione europea NextGenerationEU

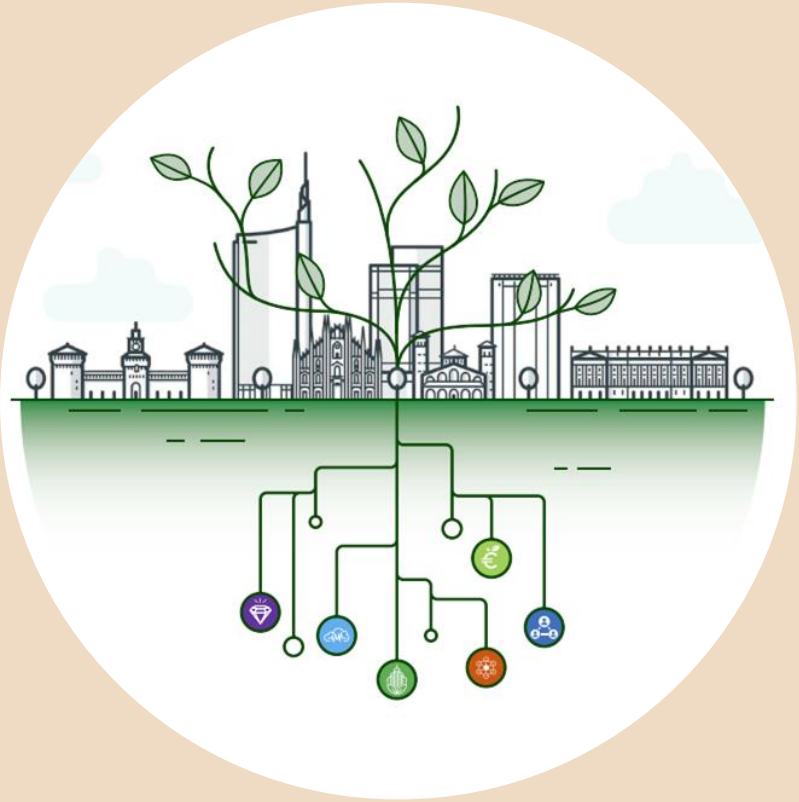
Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

MUSA

SPOKE 2
Big data
Open data
in Life Science

Claudio Ardagna



Second MUSA General Meeting

3.11.2023

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

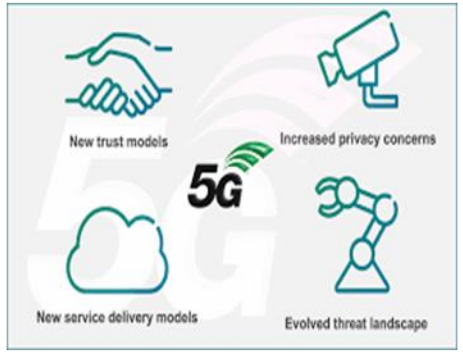
Prof. Gianvincenzo Zuccotti
Dept. of Biomedical and Clinical Sciences
Università degli Studi di Milano
co-PI of MUSA Spoke 2

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

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

Dr. Arianna Pisati
Dept. of Pharmaceutical Sciences
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 - Workpackages

WP5 – Innovation and value creation

WP2 – Cloud service platform

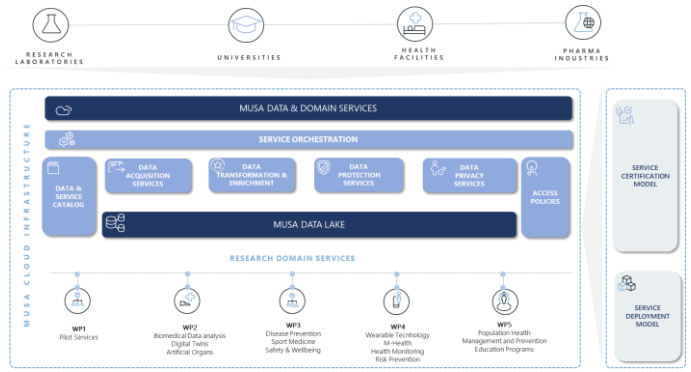
WP4 –Medical devices

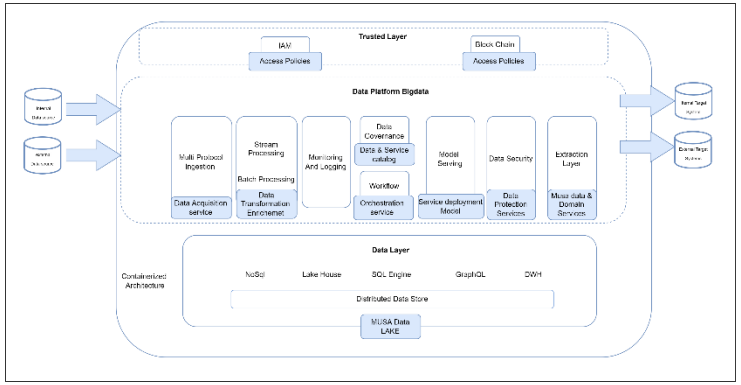
WP3 - Support healthy living

WP1 – Edge and network infrastructure

Spoke 2 – WP1 – 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
  - First working **proof of concept**



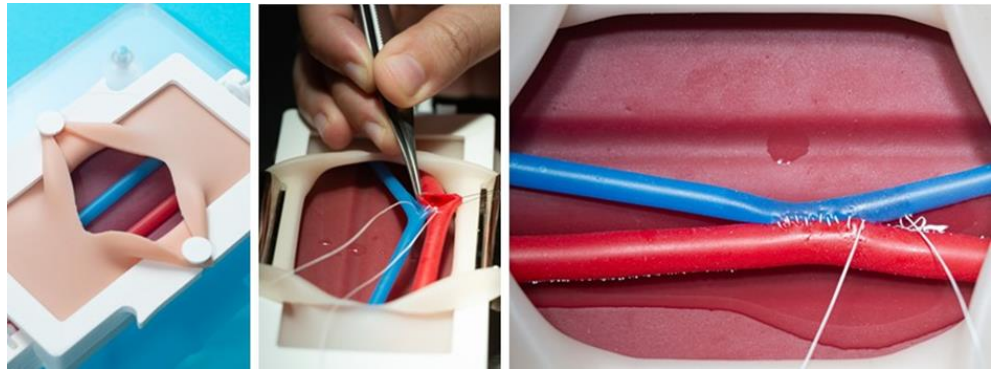


Spoke 2 – WP2 – Using Big Data for the development and sharing of new technologies in life sciences and medicine research

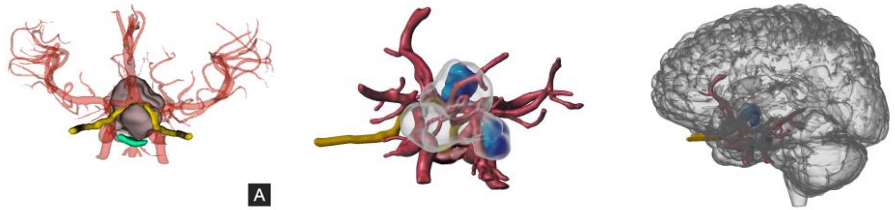
- Design and deployment of a **service platform for telemedicine** with secure collection and management of clinical data
  - Set-up of **Telemedicine Service Infrastructure**: planning actions and specs
  - Development of **Telemedicine Pilots** in collaboration with Policlinico di Milano: current focus is on two pilots, Mild Cognitive Impairment (MCI) and Heamophilia
  - Development of a multidimensional and multistakeholder framework for the analysis of **enabling factors** and the **evaluation of outcomes** (Co-production in healthcare)

Spoke 2 – WP2 – Using Big Data for the development and sharing of new technologies in life sciences and medicine research

- Study, development and production of **5 phantoms haptic prototypes** for training in medicine and surgery and for surgical planning

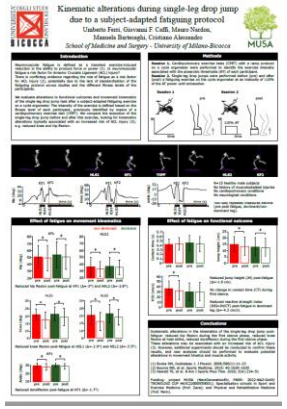


- Study, development, and creation of a **3D physical and virtual model of a real human craniopharyngioma** for surgical planning



Spoke 2 – WP3 – Big data and innovative approaches to improve global health and wellbeing

- Analysis of **behavioral Big Data** (lifestyle, sports) for disease prevention in general population
  - Study of two **protocols for myopia progression screening and cardiovascular diseases prevention**
  - Investigation of the role played by fatigue in sport injuries and new guidelines to improve the exchange of good practices among athletes, federations, and institutions
  - Two Apps for the promotions of works well-being and stress management



Spoke 2 – WP4 – Development of technologies and customizable tools for continuous monitoring, wellbeing and health

- Design and innovation on secure data collection methods and certified medical devices
  - Definition of **experimental protocols** for the validation of wearable devices for application in relevant use cases (Stress, Sport, Cardiovascular prevention, Pregnancy)
  - **Computational methods** for data analysis, feature extraction for stratification, prediction, and prevention for cardiovascular disease prevention and pregnancy monitoring
  - Identification of functional requirements and technical specification for the development of a **web-app interface for data collection and analysis**

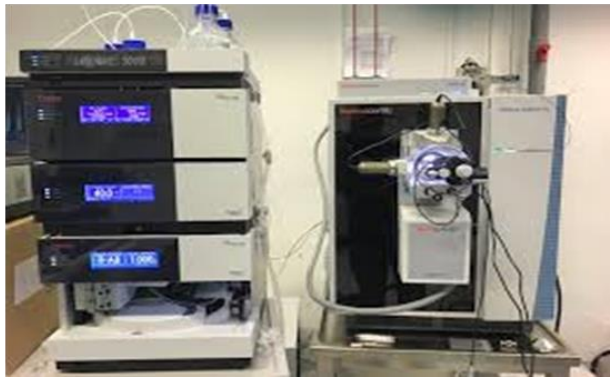
Spoke 2 – WP5 – Develop, implement, and sustain technological innovation in health

- Scouting of innovation opportunities from Spoke 2 technologies. Models to estimate value generation and sustainability of the overall infrastructure
  - **Scouting of innovative start-up projects and solutions**
  - Fostering of DHT innovation through **harmonized regulatory and assessment efforts**
    - European Taskforce for Harmonised Evaluations of Digital Medical Devices (DMDs)
    - International Digital Health Regulatory Pathways
    - National Scientific Technical Committee to support the Parliamentary intergroup on Digital Therapeutics
  - Support for **AI/ML implementation** by healthcare providers

Spoke 2 – Outlook to the future

MIND

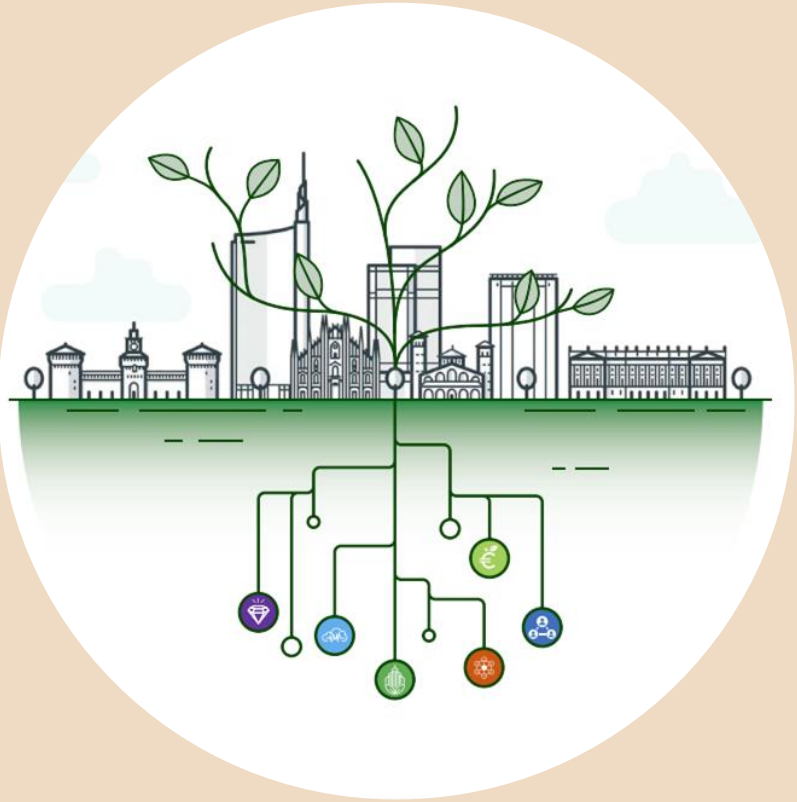
EVERYWHERE



- **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**

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

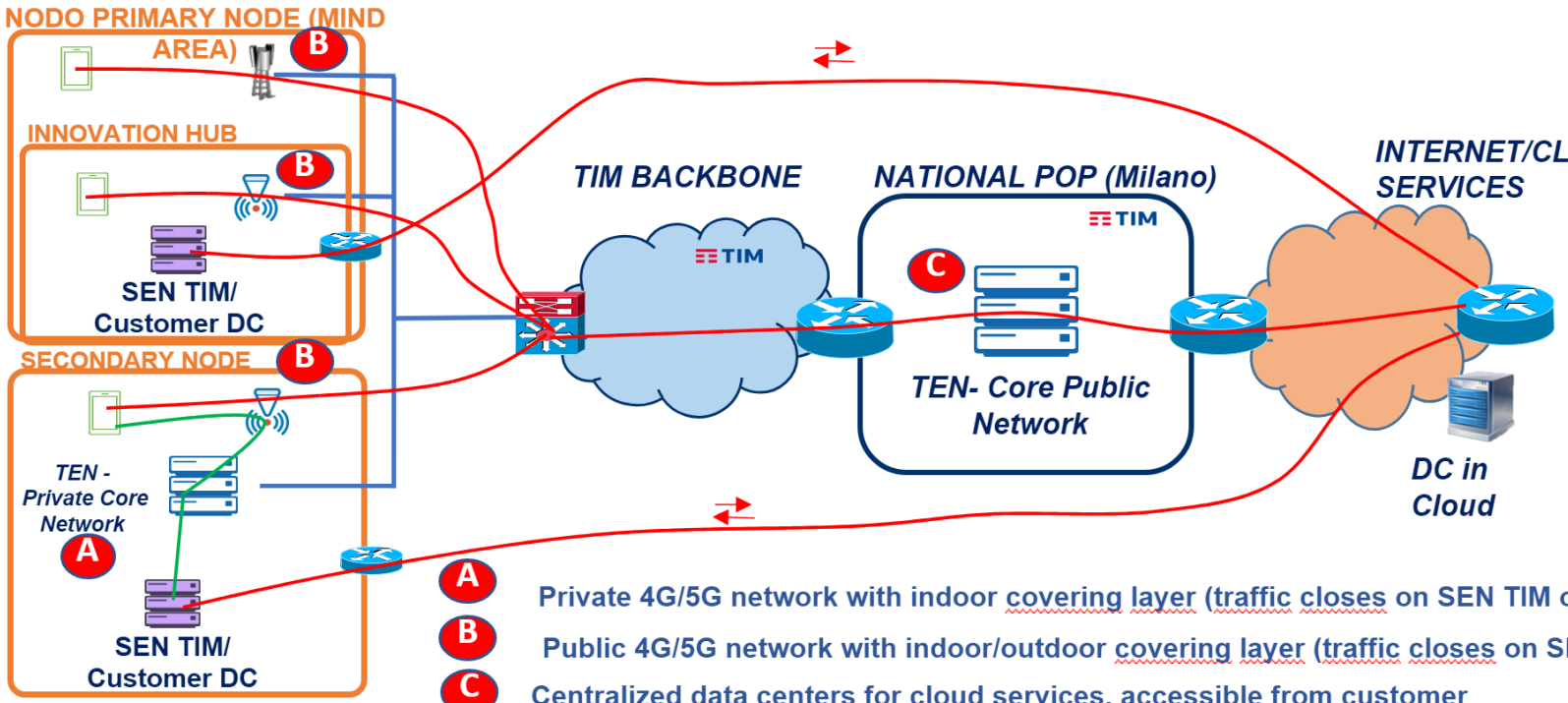
A digital platform for data analytics pipeline management in the cloud-edge continuum



Nicola Bena, UNIMI

3 November 2023

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

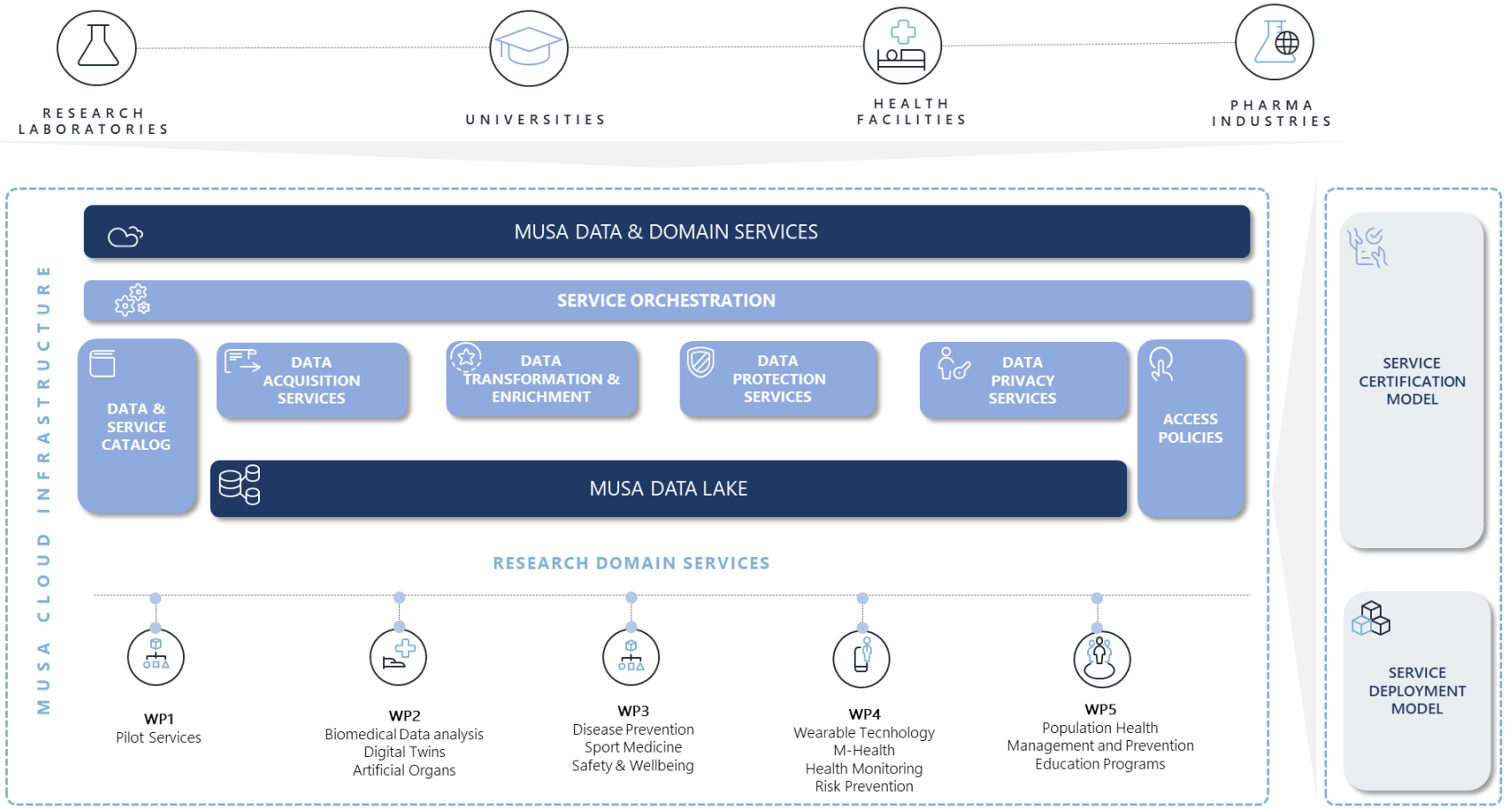
A Private 4G/5G network with indoor covering layer (traffic closes on SEN TIM or at customer-location DC)

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

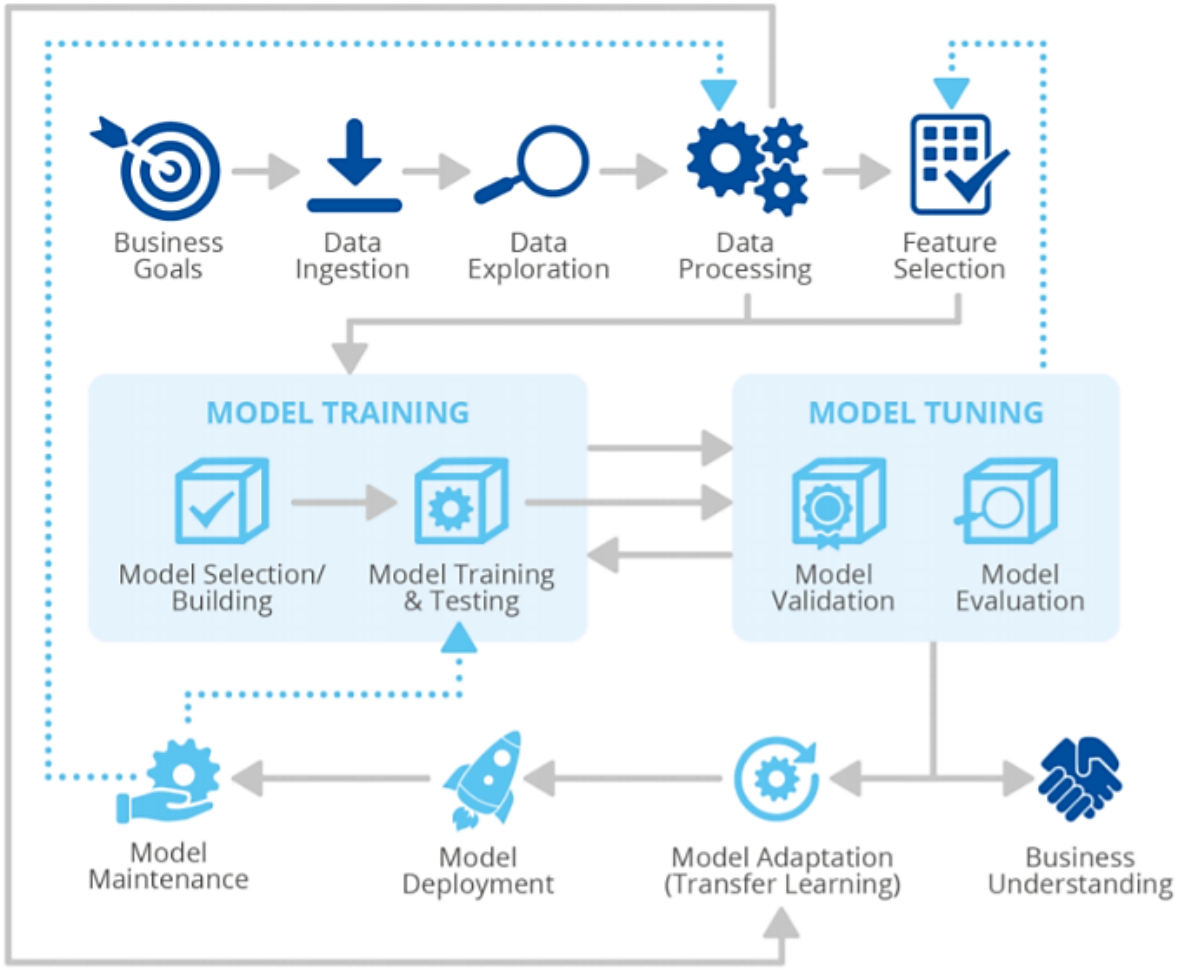
C Centralized data centers for cloud services, accessible from customer locations

Cloud Hub Data Architecture

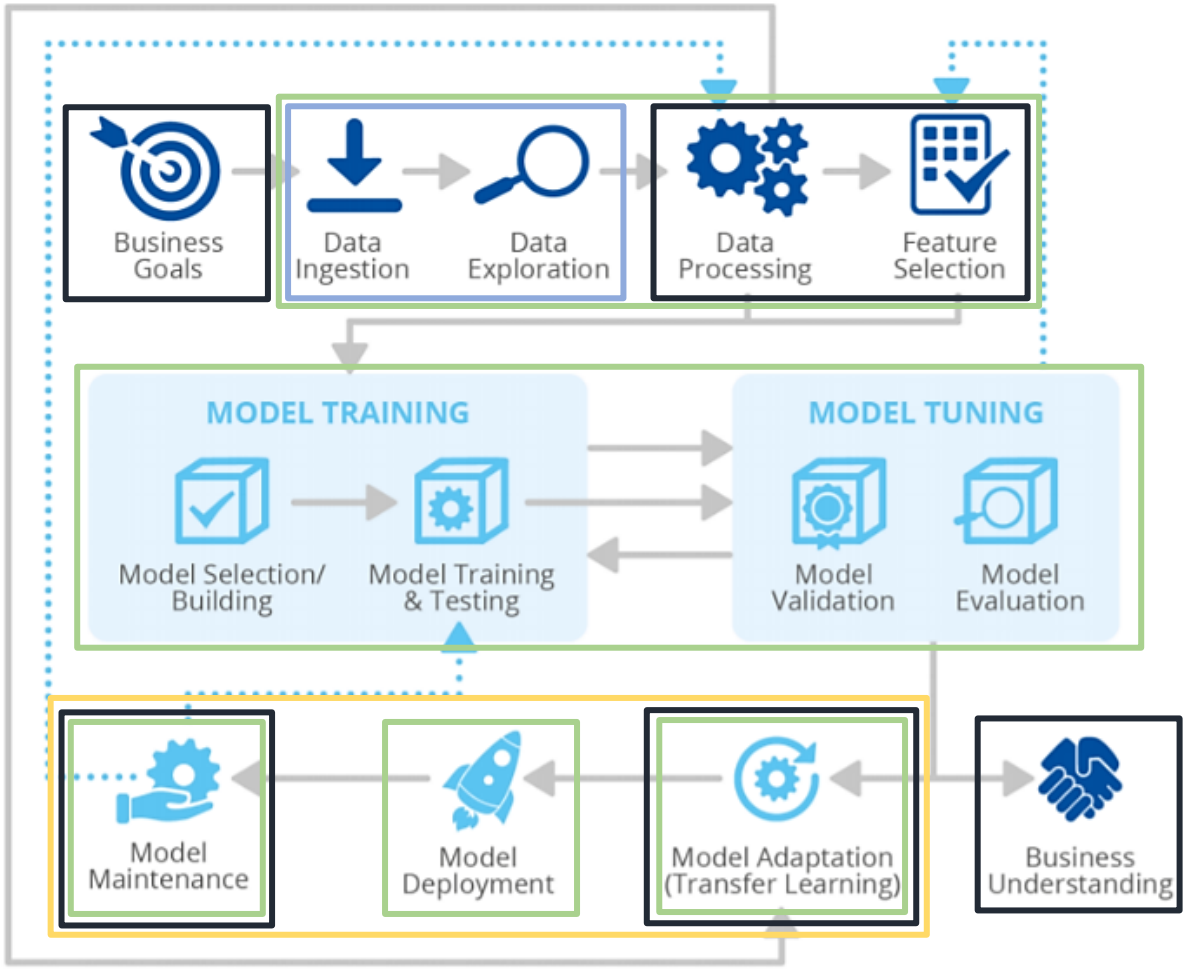
- Data lake
- Data Services
- Access Services



The complete life cycle of the data analytics platform



The complete life cycle of the data analytics platform

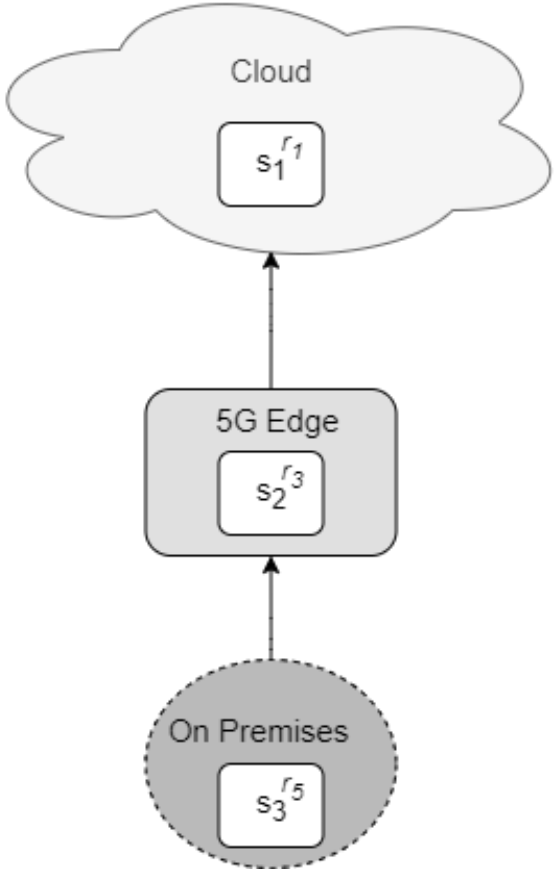


Pipeline deployment – 1

- Composition and automatic deployment of services driven by QoS requirements
- Support multiple service providers in the continuum thanks to the MUSA agents
- Containerization and monitoring based on assurance/certification

Pipeline deployment – 2

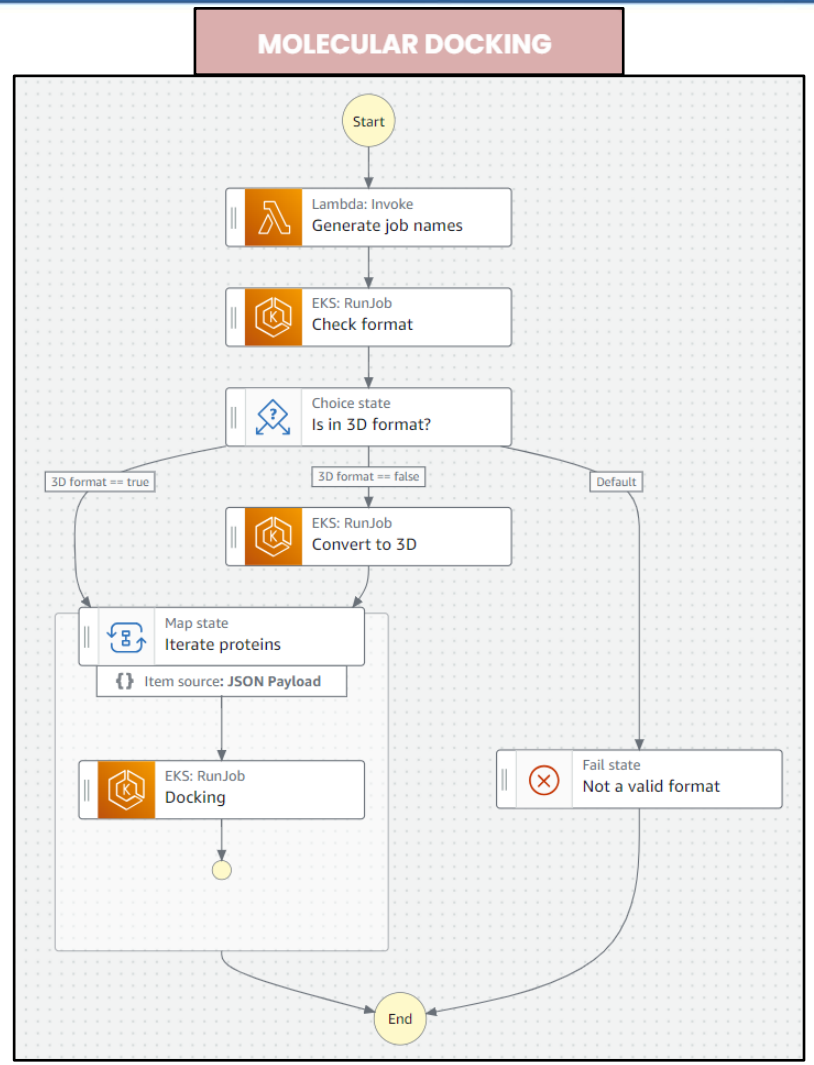
- Stakeholders use the MUSA platform to deploy their service pipelines according to QoS requirements
- CSP offer infrastructure and functionalities for deployment
- Telco operators offer 5G MEC and core network functionalities



Cloud MUSA Pilot

Analysis pipeline based on the composition of personalized and native MUSA services

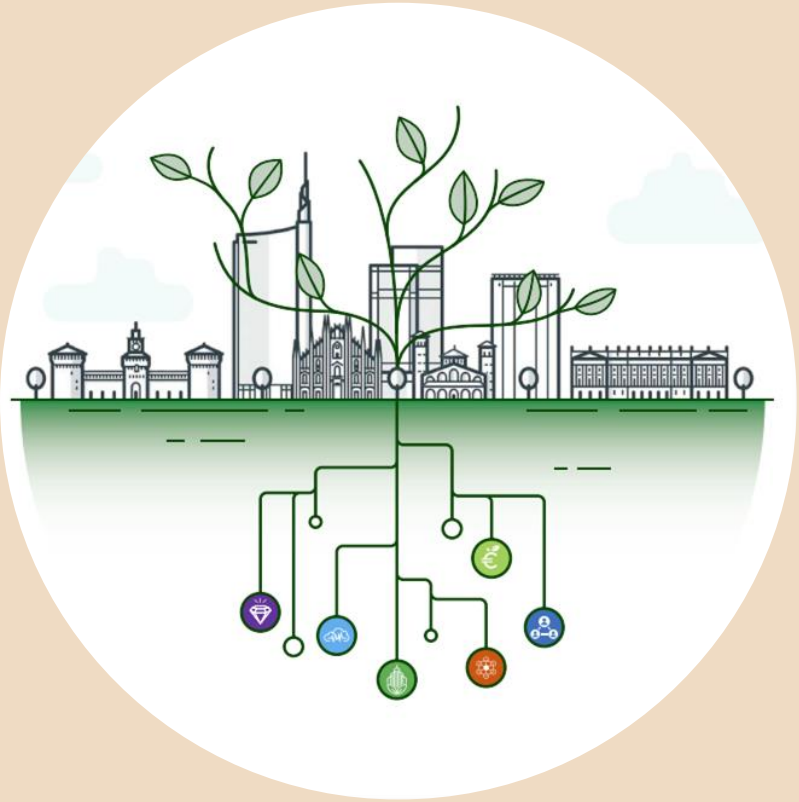
- Goal: Molecular Docking (more on this soon)
- Service containerization
- Deployment on AWS Cloud
  - Kubernetes (AWS EKS) and AWS step functions



Pilot *structure-based in silico target identification*
Guided by Prof. Giulio Vistoli
Department of Pharmaceutical Sciences, Università degli Studi di Milano

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

DELTA: Database of Enhanced Ligands and TArgets for virtual screening and activity prediction



Arianna Pisati, UNIMI

3 November 2023

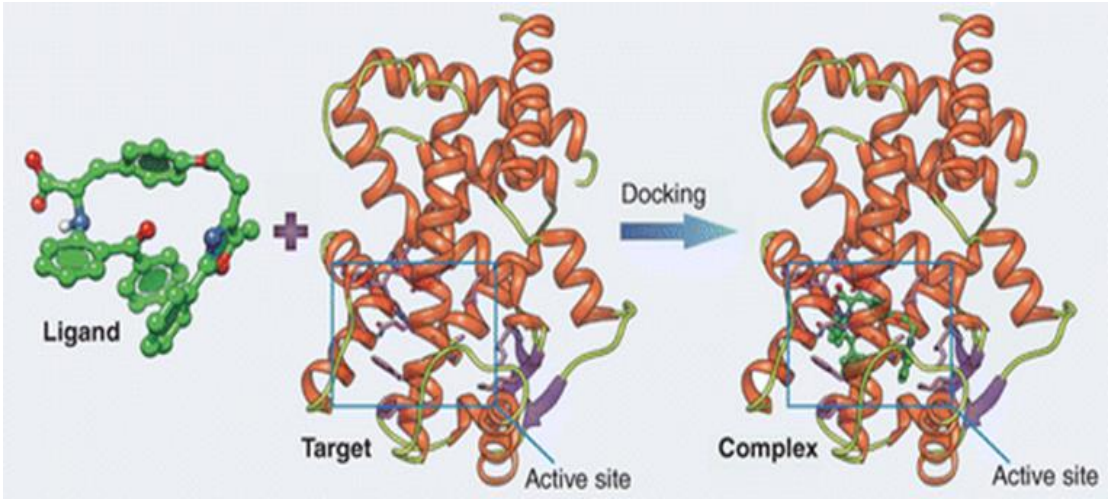
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

General introduction

Target discovery is one of the essential steps in modern drug development, and the identification of promising targets is fundamental for developing first-in-class drug.

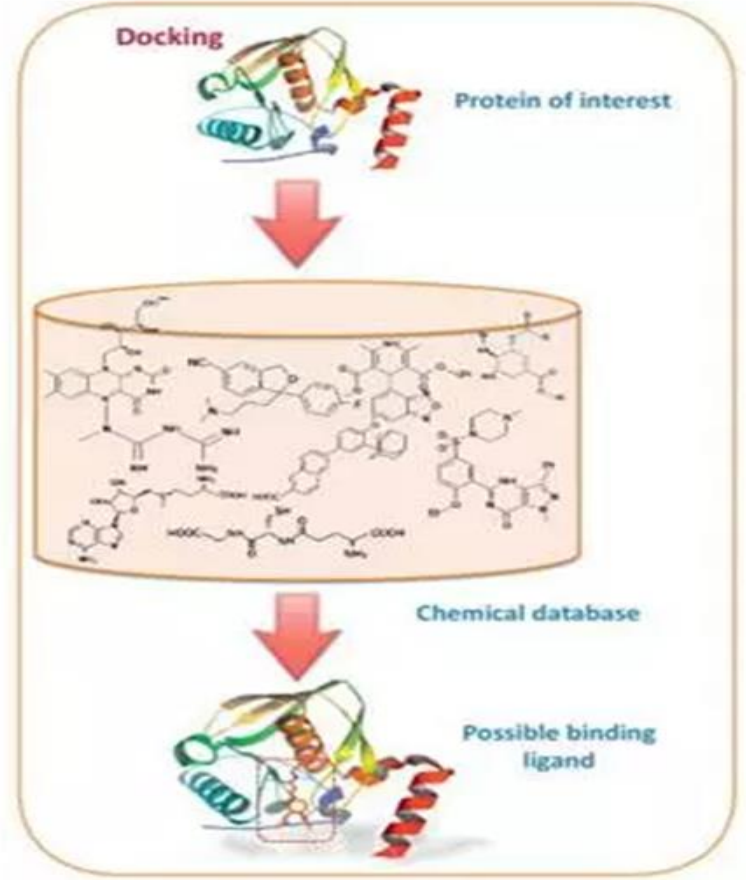
Molecular docking is a key tool in computer-assisted drug design. The goal of ligand-protein docking is to **predict the predominant binding mode(s)** of a ligand with a protein of known three-dimensional structure.

It provides a **fast, low-cost** alternative to the experimental screening of large compound libraries



Molecular docking

In molecular docking, **many small molecules** are typically docked into a **given protein** and their **binding free energies** are estimated by a value also known as "**score**" given to the docked **binding conformations** (**poses**).



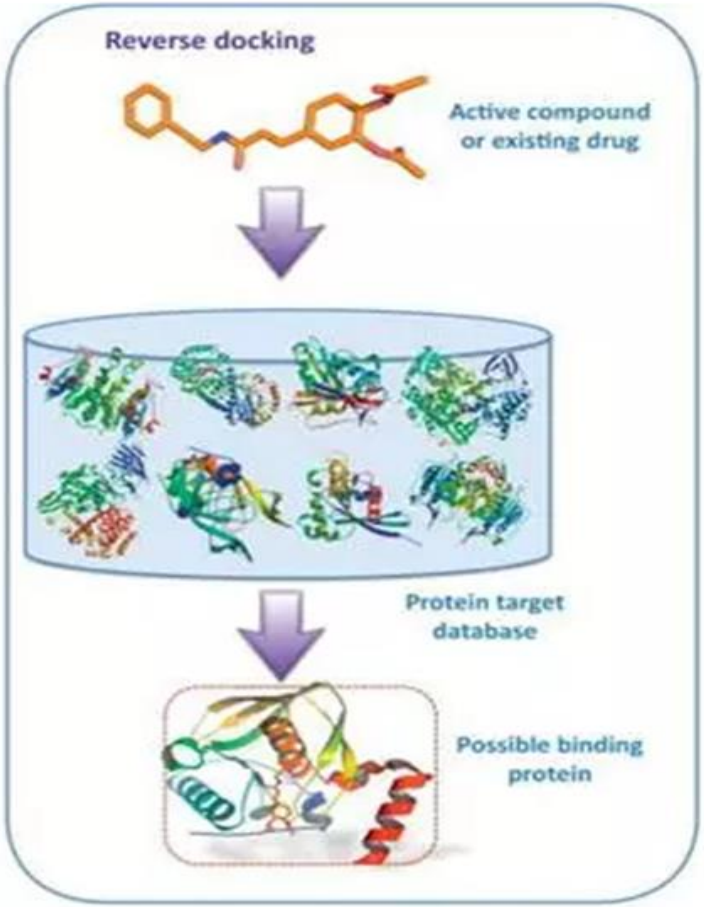
Molecular docking

However, a small molecule drug may interact with many other proteins (**off-targets**), which can have significant impacts on drug's overall biological activity, efficacy, promiscuity, and side-effects.

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

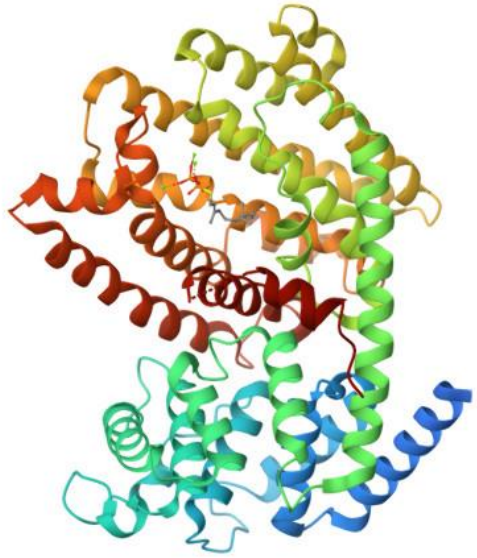
Inverse (reverse) docking

In inverse docking, a **single small molecule** is docked into a **collection of protein** structures enabling early prediction of a drugs side-effects, as well as toxicity. Inverse docking therefore plays an important role in modern drug discovery and design.



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

Targets database



DELTA currently contains **485 protein structures**, which have been collected based on three main criteria:

Therapeutically relevant

- It is associated with a particular disease and may be inhibited or activated by a therapy in a way that will change the course of the disease in a positive way

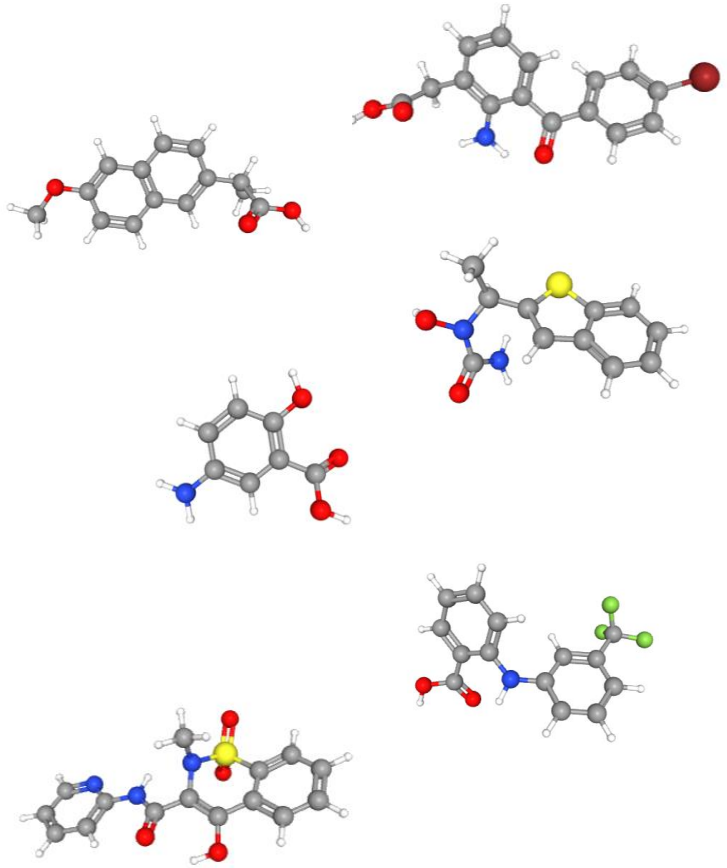
Available drug

- They are targeted by at least a drug (either approved or in clinical trial phase)

Available ligands

- At least 100 active and 100 inactive compounds with experimentally determined activity are available in litterature

Ligands database



DELTA currently contains **113061 ligand structures** which, for each target, have been collected based on four main criteria:

Organic molecules

- We discarded small inorganic compounds and ions

Molecular weight

- We only selected compounds with MW<1000

Activity

- We defined an activity threshold to disciminate between active and inactive compounds

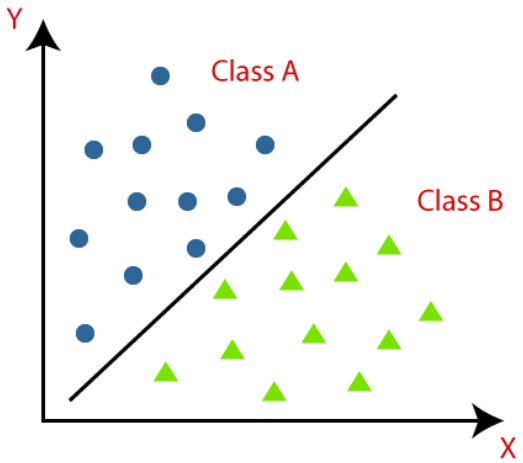
Structural diversity

- A cluster analysis has been performed to select 200 compounds (100 active and 100 inactive) with structure as diverse as possibile

DELTA's Applications

- Development of **local classification models** based (for each target) on the **sole ligand features**
- Development of **local classification models** based (for each target) on **docking-based features**

The **combination** of ligand- and docking-based descriptors will be also evaluated to **enhance the predictive power** of the models

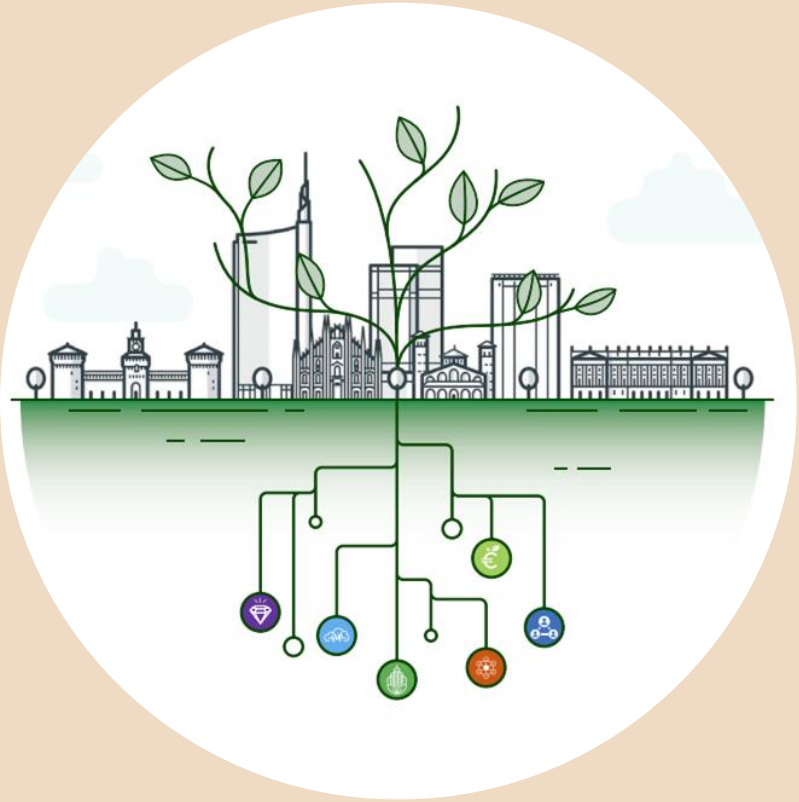


The **purpose** of such models is to **predict the activity of a given ligand on each protein**

Spoke 2

DEMO

3 November 2023



Finanziato dall'Unione europea
NextGenerationEU

MUSA

DEMO

MUSA WP1 Architecture

Ernesto Damiani, Claudio Ardagna, Marco Anisetti, Ruslan Bondaruc