

Progetti R&D a supporto delle attività di AI

Giacinto Donvito
INFN-Bari
(donvito@infn.it)

Agenda



- Contesto del talk
- Progetti conclusi o in corso
 - DEEPHybrid-DataCloud
 - CNRBionics
 - IoTwins
 - PLANET
 - health big data
- Progetti approvati in partenza:
 - InterTwin
 - AI4INFN
- Proposal sottomessi in attesa di approvazione
- Conclusioni e considerazioni

- Cercherò di raccontare le attività in corso in progetti di R&D (finanziati) nell'ambito del computing,
- che hanno come core il supporto agli strumenti e algoritmi di AI/ML/DL
- cercando di dare risalto a come queste attività possano essere molto utili nel contesto della bio-medicina

DEEP-HybridDataCloud

Start date 1 November 2017

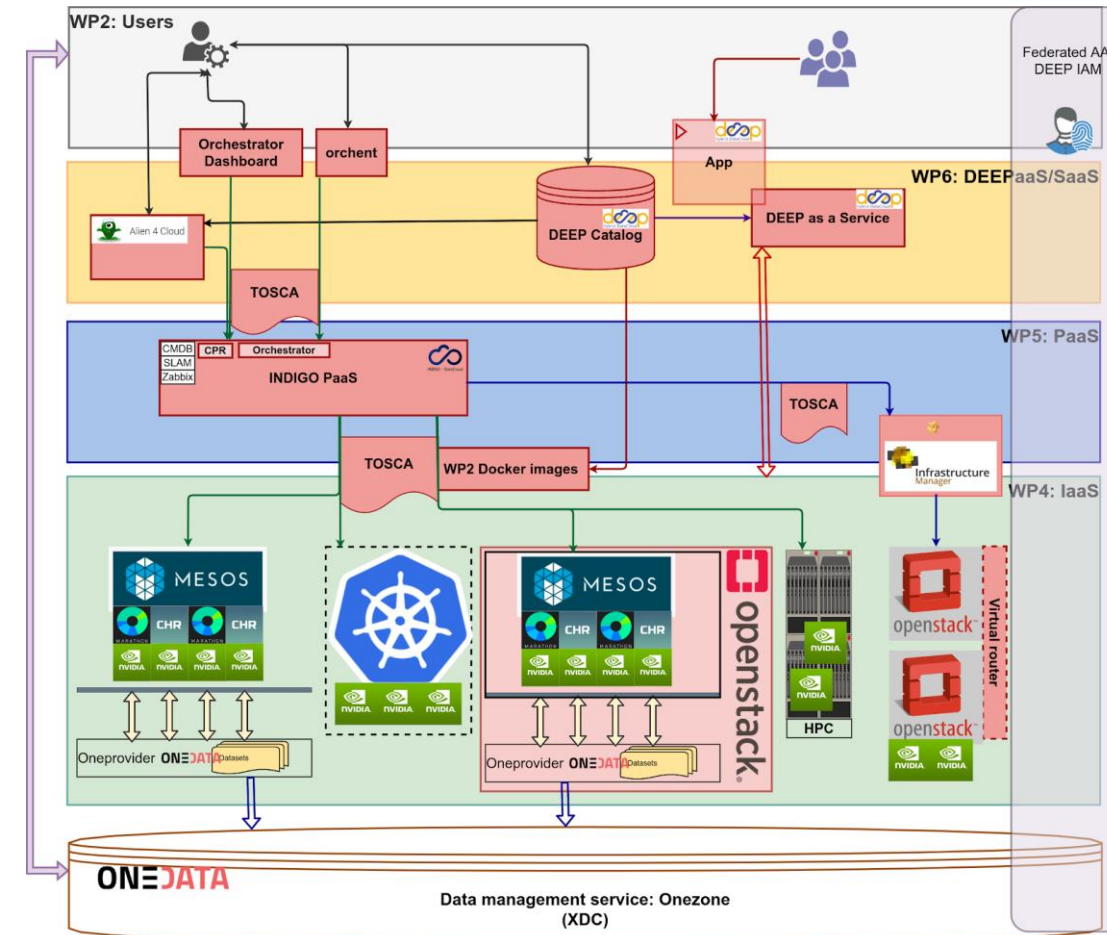
End date 30 April 2020

Grant agreement ID: 777435

H2020-EU.1.4.1.3.



- R&D partito dagli output di INDIGO-DataCloud
- Global objective: Promote the use of intensive computing services by different research communities and areas, and the support by the corresponding e-Infrastructure providers and open source projects
 - Focusing on Machine learning, Deep learning, and Post processing



DEEP-HybridDataCloud

Start date 1 November 2017
End date 30 April 2020
Grant agreement ID: 777435
H2020-EU.1.4.1.3.



- Three techniques of wide interest, involving
 - Large, heterogeneous data sets
 - Intensive computing demands that would benefit from using hardware accelerators (GPUs, low latency interconnects)
- Use cases:
 - Diabetic rethinopathy detection, biodiversity applications.
 - Objective: Provide a general, distributed architecture and pipeline to train, retrain and use deep learning (and other machine learning) models
 - Online analysis of data streams: intrusion detection systems, anomaly detection
 - Provide an architecture able to analyze massive on-line data streams, also with historical records

DEEP-HybridDataCloud

Start date 1 November 2017

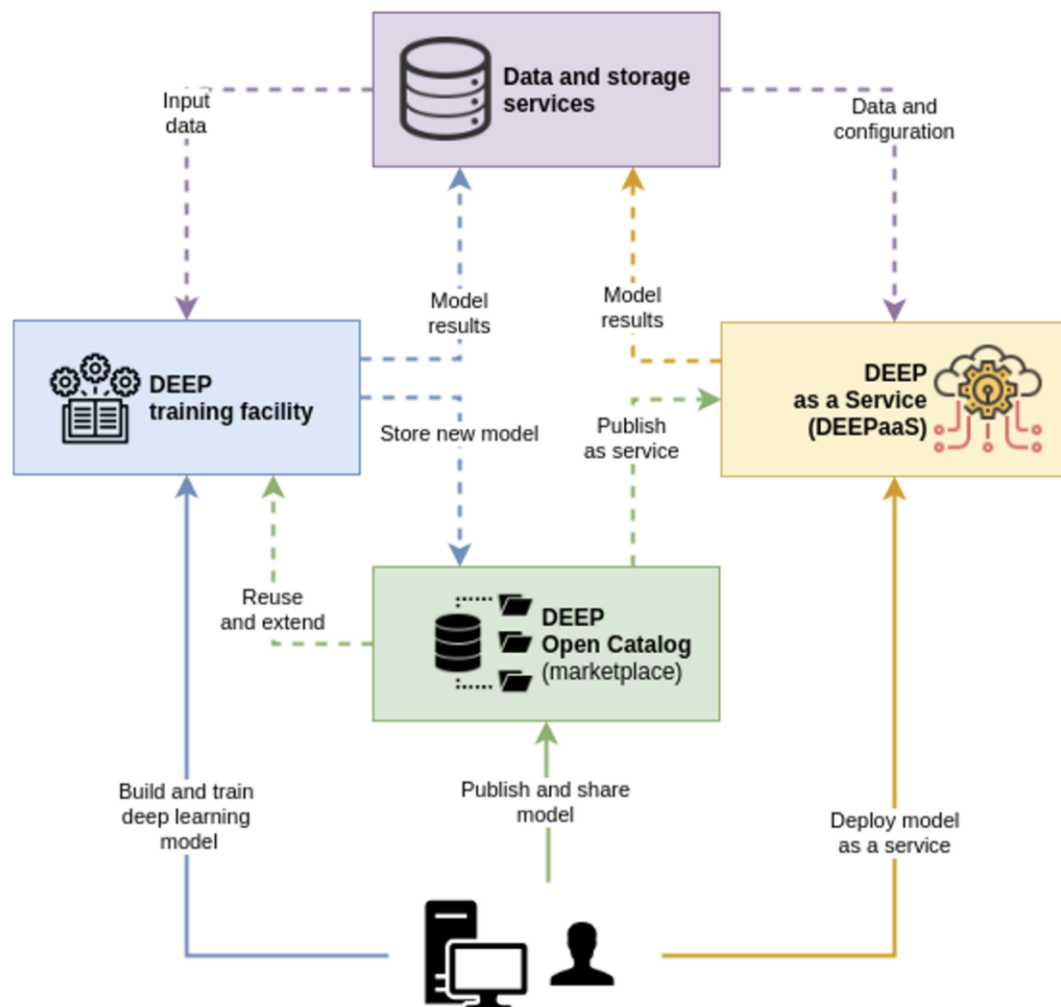
End date 30 April 2020

Grant agreement ID: 777435

H2020-EU.1.4.1.3.



High level architecture



- Position as technology providers to support DL/ML in the EOSC
- Generic building blocks (services) for exploitation through EOSC
 - **DEEP training facility**
 - **DEEP as a Service (DEEPaaS)**
 - **DEEP Open Catalog**
- Integration with storage from external initiatives (eXtreme-DataCloud, EGI Data Hub)

DEEP-HybridDataCloud

Start date 1 November 2017

End date 30 April 2020

Grant agreement ID: 777435

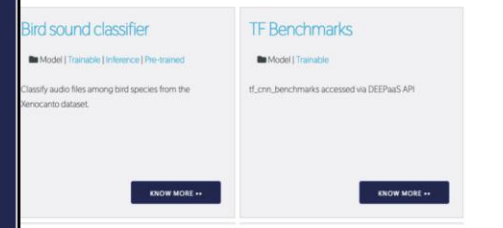
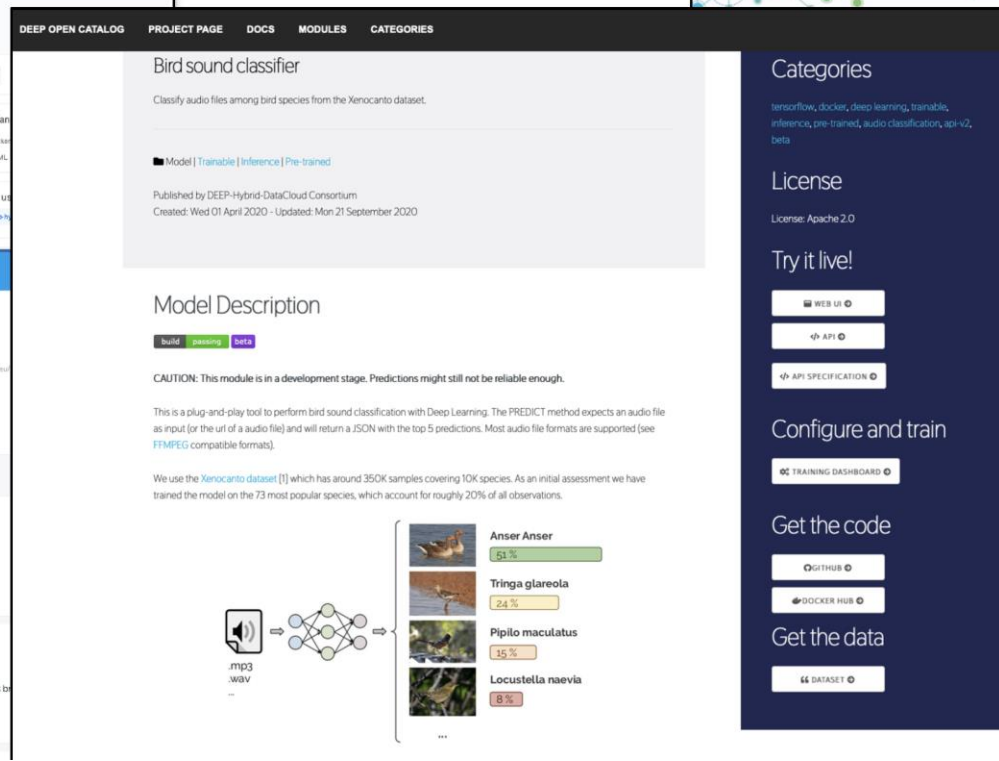
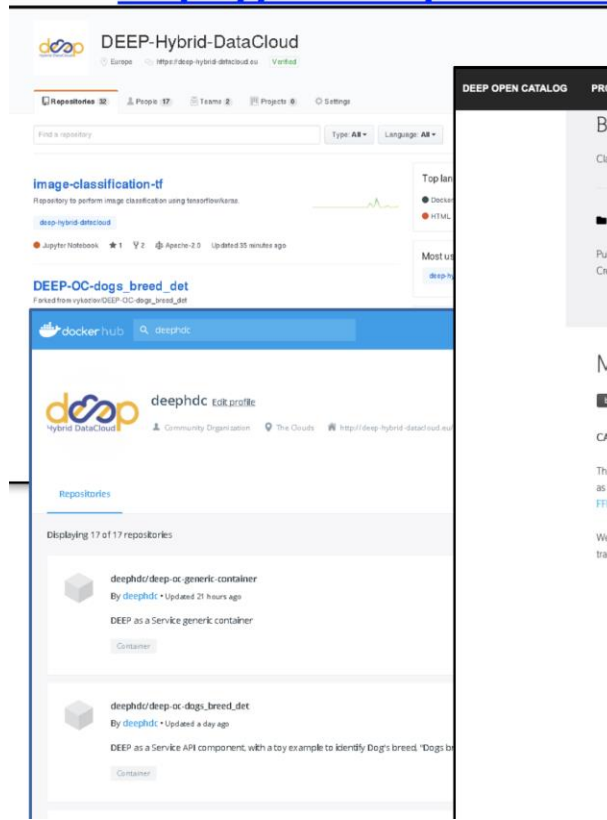
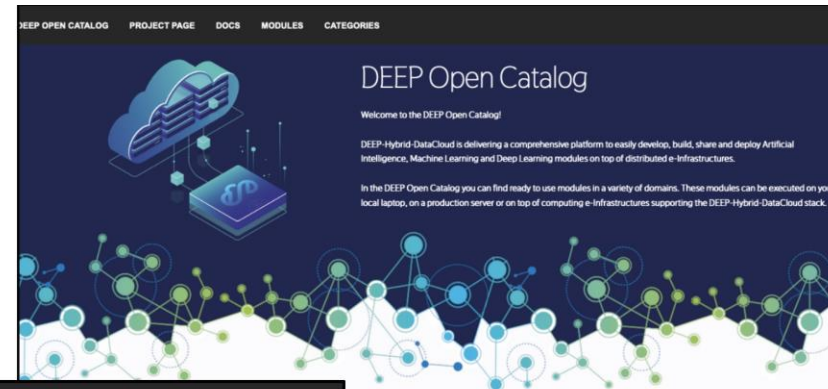
H2020-EU.1.4.1.3.



DEEP Open Catalog

Sharing knowledge between users

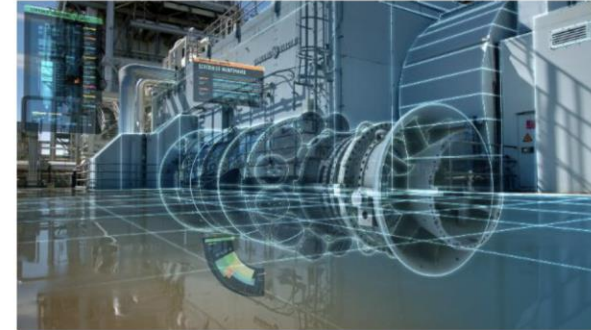
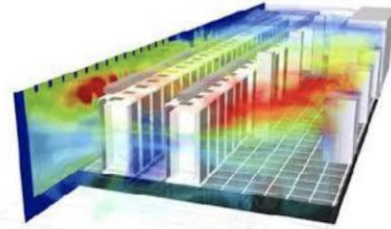
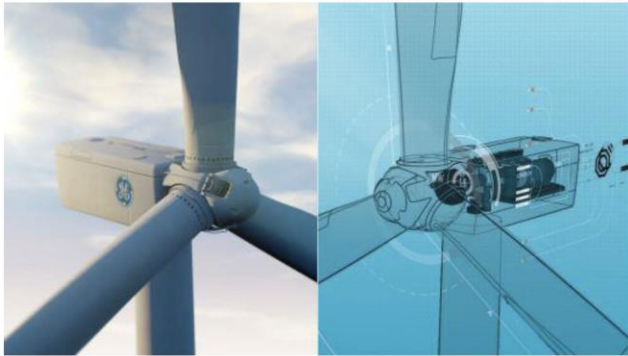
<https://marketplace.deep-hybrid-datacloud.eu>



IoTwins








Daniele Cesini

Digital Twins



- A **digital twin** is a digital replica of a living or non-living physical entity
- Digital twins integrate **internet of things**, **artificial intelligence**, **machine learning** and **software analytics** with spatial network graphs to create **living digital simulation models** that update and change as their physical counterparts change
- Particularly useful to detect, diagnose and predict anomalies

Testbeds

- 1) RI-KKWS – Wind turbine predictive maintenance 
- 2) FILL-TTT – Machine tool spindle predictive behaviour
- 3) Predictive maintenance for a crankshaft manufacturing system 
- 4) GCL - Predictive maintenance and production optimization for closure manufacturing 
- 5) NOUCAMPNOU – Sport facility management and maintenance
- 6) Examon (CINECA) – Holistic supercomputer facility management 
- 7) SAG-SAGOE - Smart Grid facility management for power quality monitoring
- 8) Patterns for smart manufacturing for SMEs
- 9) Examon replication to INFN/BSC datacentres 
- 10) Standardization/homogenization of manufacturing performance 
- 11) FCB - Replicability towards smaller scale sport facilities
- 12) MARP - Innovative business models for IoTwin PaaS in manufacturing 

IoTwinS

IoTwinS Architecture

- The primary elements of the IoTwinS PaaS infrastructure at the cloud side will be provided using state-of-the-art technologies, such those developed in H2020 initiatives like the INDIGO-, XDC-, and DEEP-DataCloud projects
- At the PaaS level, INDIGO and DEEP provide solutions for automatically instantiate and orchestrate virtual resources to run ML applications as soon as new data appears.
- Storage events notifications are used to trigger new models training or new analysis/inference.
- Multiple Authorization and Authentication mechanisms are also provided, based on X509 certificates, tokens (openIDConnect) or plain username/ password.

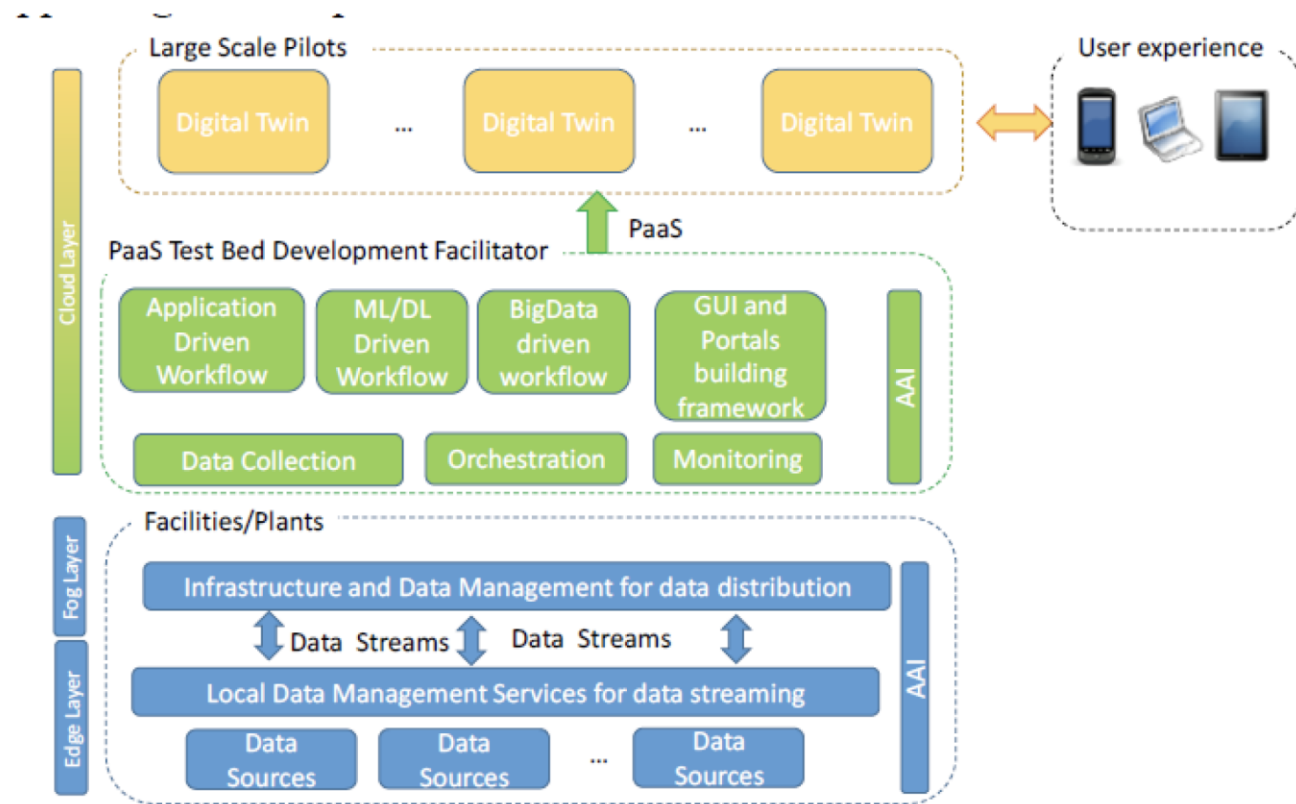


Figure 5: Cloud-based IoTwinS infrastructure

Implementation - 1

WP No	Work Package Title	Lead Part. No	Lead Part. acronym	PM	Start M	End M
1	Project management	1	BRI	80	1	36
2	IoT-Edge-Cloud infrastructure and big data services for SMEs	4	FOKUS	493	1	36
3	AI services for distributed digital twins	8	UNIBO	153	1	30
4	Manufacturing test-beds	17	GCL	580	1	36
5	Facility management test-beds	6	FCB	263	1	36
6	Platform replicability, scalability, and business models	3	ESI	258	1	36
7	Exploitation of the project developments and knowledge management	12	WST	181	1	36
8	Outreach activities – dissemination and communication	11	ASTER	150	1	36
				2158		

• INFN

- WP1: 1PM
- WP2: 61PM
 - Leader of **T2.1: Technology Requirements**
 - Leader of **Task 2.6: Service Orchestration and Security**
 - Leader of **Task 2.8: Cloud-side Data Services Development**
 - management of massive streams of data in order to analyse them as they are produced, so as to exploit their "freshness", by possibly using approximate techniques for low power computing and/or for reducing time computation
 - data management services will provide capabilities for managing the entire data lifecycle, allowing policy definition to control it. Policies will allow to request and enforce attributes such as:
 - » replica multiplicity
 - » storage Quality-of-Service (i.e. fast vs glacier-like resources)
 - » access control evolution over time (i.e. open availability after "embargo" periods characterized by limited access).
 - The data management system will integrate
 - the encrypting, anonymizing, and ACL enforcing

PLANET project: Open-source and cloud-native solutions for managing and analyzing heterogeneous and sensitive clinical Data

Daniele Spiga

The PLANET Project in a nutshell

PLANET (**P**ollution **L**ake **A**nalysis for **E**ffective **T**herapy) a INFN-funded research initiative aiming at developing

- An observational (ecological) study **to evaluate the association between air pollution and Covid19, taking care of a variety of components that are supposed to influence rates of SARS-COV-2 diffusion and infection**
 - A synergy between INFN and epidemiological and medical knowledge University of Perugia

The study is based on the hypothesis that pollution may contribute to the spread and/or the severity of COVID-19, through 2 possible mechanisms:

- **Acute:** microparticles derived from fossil fuels might act as airborne carriers of virus;
- **Chronic:** exposure to microparticles and chemical pollutants might cause chronic lung injury, exacerbating the consequences of viral infection.

The PLANET Project in a nutshell

The Analysis Strategy

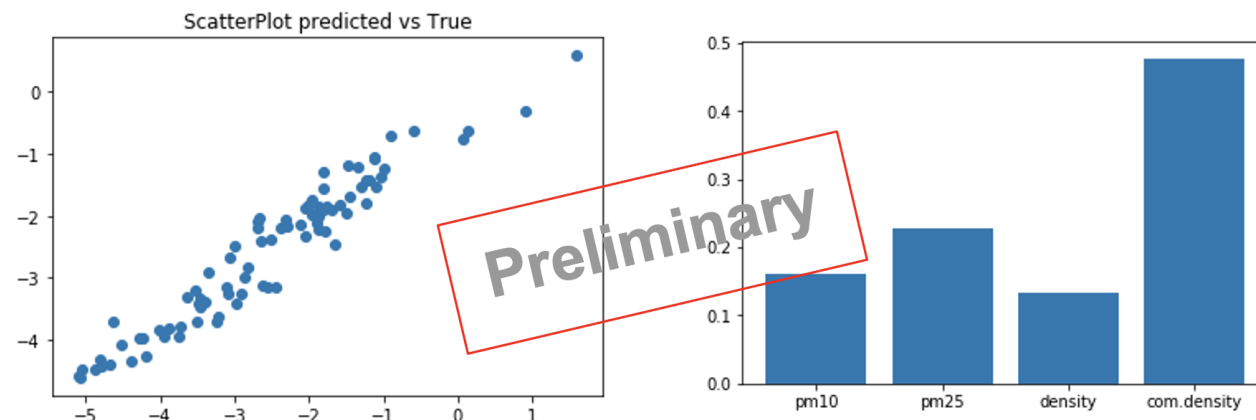
A key element of PLANET is to include a wide variety of components that are expected to influence rates of SARS-COV-2 diffusion and infection

- atmospheric data, population density, urban vs rural environment, mobility, socio-economic conditions...
- take into account also the fact that severity COVID-19 disease and deaths are also influenced by many variables (age, gender, comorbidities, frailty, etc..).

Current focus on Feature importance

evaluation: assign a score to input features based on how useful they are at predicting a target variable (Covid19)

- Models such as: Random Forest; k-nearest neighbors;



The PLANET Project in a nutshell



INFN key expertise

INFN is a pioneer in the **design and implementation of large-scale computing infrastructures and applications**

- Primarily developed to meet the needs of the latest generations of high energy physics (HEP) experiments
- **now rapidly extending to other communities. INFN@PLANET:**
 - [ML INFN](#); [CYGNO](#); [INFN-Cloud](#)

• **Data Analysis**

To develop models and to implement statistical data analysis

• **Data Curation and Data Management**

Organization and integration of data collected from heterogeneous sources;
Enabling FAIR (Findable, Accessible, Interoperable, Reusable) data repositories

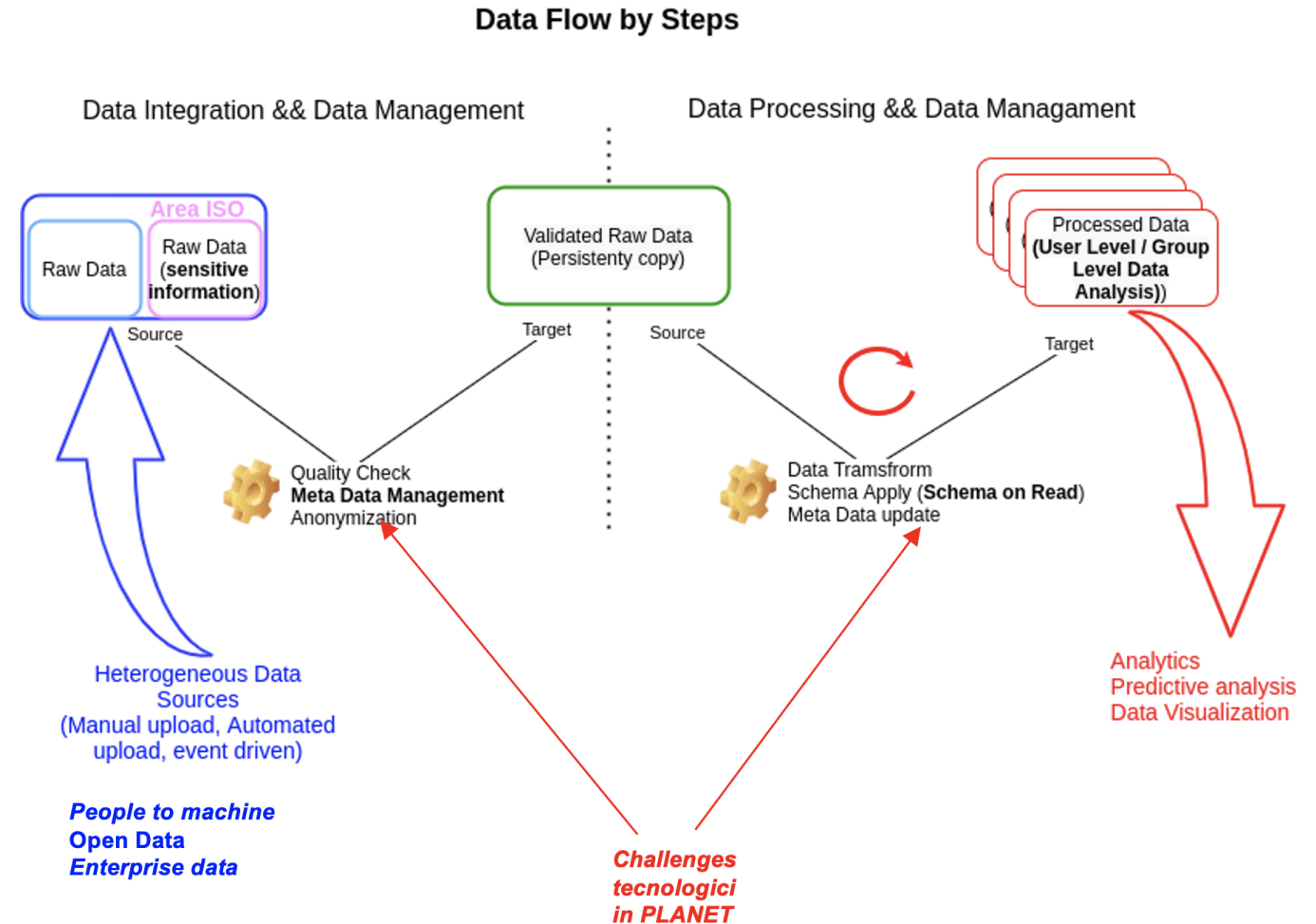
• **Integrated and Certified Computing Infrastructure**

A ISO 27001 / 27017 Certified Data-Lake to manage confidential data (ISS, Hospital, ASL)
An-easy-to-use computing platform fully integrated with the **INFN-Cloud national infrastructure**

The PLANET Project in a nutshell

Toward a Data Lake model

- **A data storage environment that allow to keep data in their native format.**
 - They remain in this condition until it is necessary to define a structure
- **Enable automated data validation (and organization)**
- **Data scheme is defined at the time of analysis and not at the time of archiving**



health-big-data

Davide Salomoni

- **Abstract**

- The ten-year HBD project, funded by the MEF and coordinated by the Ministry of Health, involves 51 IRCCS belonging to the Networks Alliance Against Cancer, Neuroscience and Neurorehabilitation and Cardiology managed by the three Networks in collaboration with the Politecnico of Milan.
- The project provides for the creation or enhancement of: i) a set of local IT platforms in each participating IRCCS to ensure extraction, integration and interoperability of clinical and scientific data; ii) a centralized IT platform, ensuring connectivity between IRCCS and advanced analysis of shared data. The type of data that will be collected and shared is heterogeneous and includes omics data (genomics, transcriptomics, proteomics, metabolomics), clinical data (electronic medical record and patient follow-up data), clinical imaging and radionics data; and data provided by the patient. Data from biosensors, environmental, social, and economic data will also be included in the medium term.
- The platform will ensure connectivity between the IRCCS participants in the project, other Italian and non-Italian Research Institutes, the databases of the National Health Service and international public databases, to develop predictive and prescriptive analysis capabilities based on the integration of omics and clinical data and the study of aggregate patient subgroups for various biological and clinical parameters, with the possibility of access and sharing by individual operators of each IRCCS. Besides this, the platform will provide essential operational tools to the Molecular Tumor Boards, which have already been initiated in several network IRCCS.

health-big-data

- Collected data:
- The type of data that will be collected and shared is heterogeneous and includes:
 - Omics data
 - Genomics, transcriptomics, proteomics, metabolomics
 - Clinical data
 - Electronic medical records and patient follow-up data
 - Imaging data
 - Clinical and radiomics imaging data
 - Individual data
 - Data provided by the patient

health-big-data



- Il team di progetto:
 - Rete IRCCS delle Neuroscienze e della Neuroriabilitazione
 - Rete Cardiologica degli IRCCS
 - Politecnico di Milano
 - Ministero della Salute
 - INFN Italian National Institute for Nuclear Physics

Attività approvate che partiranno a breve

interTwin

Daniele Spiga

The Project in numbers

CALL: **HORIZON-INFRA-2021-TECH-01-01** (submitted on **23 September 2021**)

- Next generation of scientific instrumentation, tools and methods

INFN ROLE: Partner

SCIENTIFIC AREA: Innovative and customizable services for EOSC

PROJECT NAME: **An interdisciplinary Digital Twin Engine for science (interTwin)**

PROJECT COORDINATOR: EGI Foundation (31 Partners)

DURATION: 36 Months

INFN DEPARTMENT: **Perugia, CNAF, Bari, Torino, Pisa**

- National Coordinator: Daniele Spiga

BUDGET:

- **Total Budget: 12.000.000 €**
- **Budget for INFN: 1.029.375 €**

PROJECT STARTS: 01.09.2022

Main objectives of the project

Objective 1. Co-design, develop and provide a Digital Twin Engine that simplifies & accelerates the development of complex application-specific DTs that benefits researchers, business and civil society.

Objective 2. Co-design a Digital Twin Engine blueprint architecture that provides a conceptual framework for the development of DTs supporting interoperability, performance, portability, & accuracy.

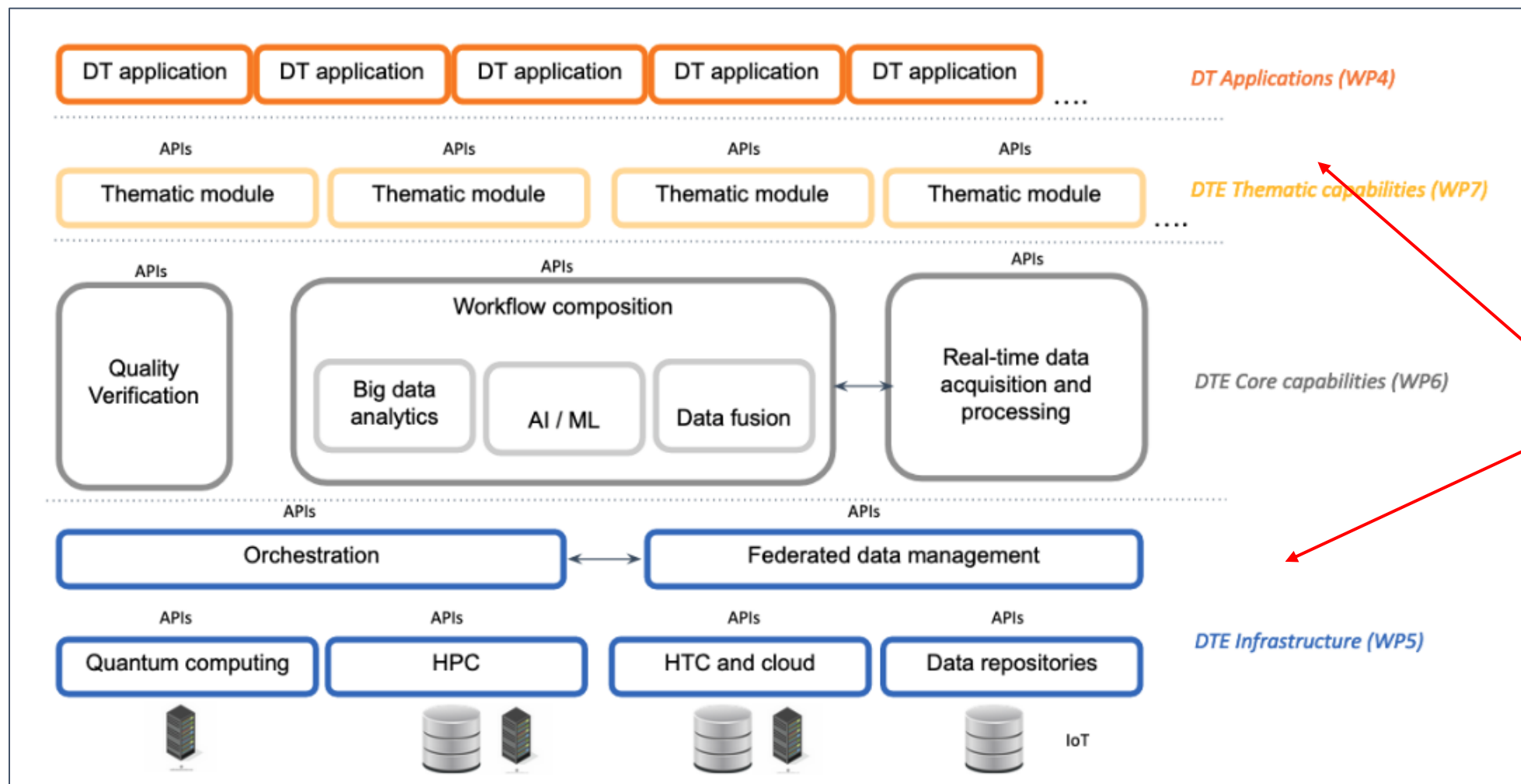
Objective 3. Extend the technical capabilities of the European Open Science Cloud with modelling & simulation tools integrated with its compute platform

Objective 4. Ensure trust and reproducibility in science through quality, reliability, & verifiability of the outputs of Digital Twins

Objective 5. Demonstrate data fusion with complex modelling & prediction technologies

Objective 6. Simplify DT application development with tools to manage AI workflows and the model lifecycle while reinforcing open science practices

interTwin Conceptual Model



(See later)

INFN Roles and responsibilities: two assets

Although INFN will contribute to several activities in the interTwin project. Main roles will be on **WP5 (Lead)** and **WP7 (Virgo/ET)**

The DTE infrastructure provides the federated data and compute resources involved in modelling and simulation

Federated data management (enhance the DataLake)

- manage HPC data ingress/egress, data transfer and use an event-based interface for event-driven analysis frameworks.

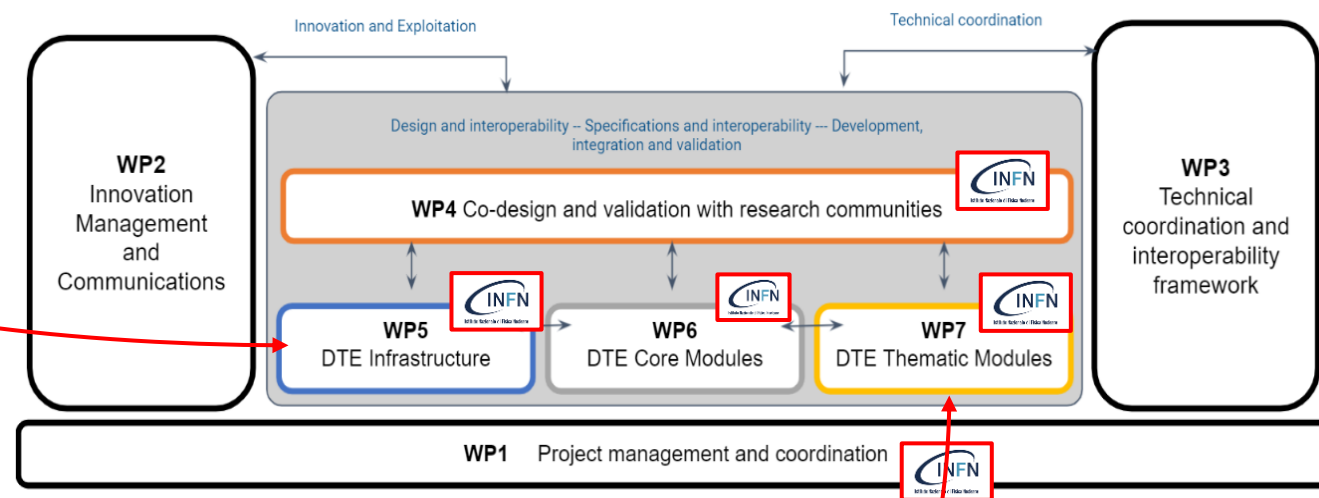
Federated Compute Infrastructure

- delivers on-demand processing capacity and federated HTC HPC etc to support both batch workflow execution and interactive analysis.

Orchestration:

- executes complex tasks by matching the best storage and compute resources, orchestrate data transfer tasks before launching the analysis.

Federation services and Authentication & Authorisation Infrastructure (AAI)



DT of the Virgo Interferometer Use Case

- The sensitivity of GW interferometers is limited by noise; its reduction and subtraction are one of the most important and challenging activities in GW research.

Develop a DT of an interferometer

- to realistically simulate the noise in the detector, in order to study how it reacts to external disturbances and, in the perspective of the Einstein Telescope, to be able to detect noise "glitches" in quasi-real time, which is currently not possible. This will allow the low-latency search pipelines to veto or de-noise the signal, sending out more reliable triggers to observatories for multi-messenger astronomy

AI4EOSC

Giacinto Donvito

- Call:
 - HORIZON-INFRA-2021-EOSC-01-04
 - Project results are expected to contribute to all the following expected outcomes:
 - increased service offer and capabilities beyond the present landscape in addressing the current and anticipated needs of the research community at large;
 - increased availability of (pre)operational services that can be customised and integrated in the existing workflows of researchers across different disciplines, facilitating the cross-disciplinary collaboration, reducing the time to results and increasing productivity.
- Scope:
 - The aim is to provide researchers with a set of highly innovative new services that would exploit, in a structural way, cloud-based EOSC technologies and European compute and data management capacities.

AI4EOSC

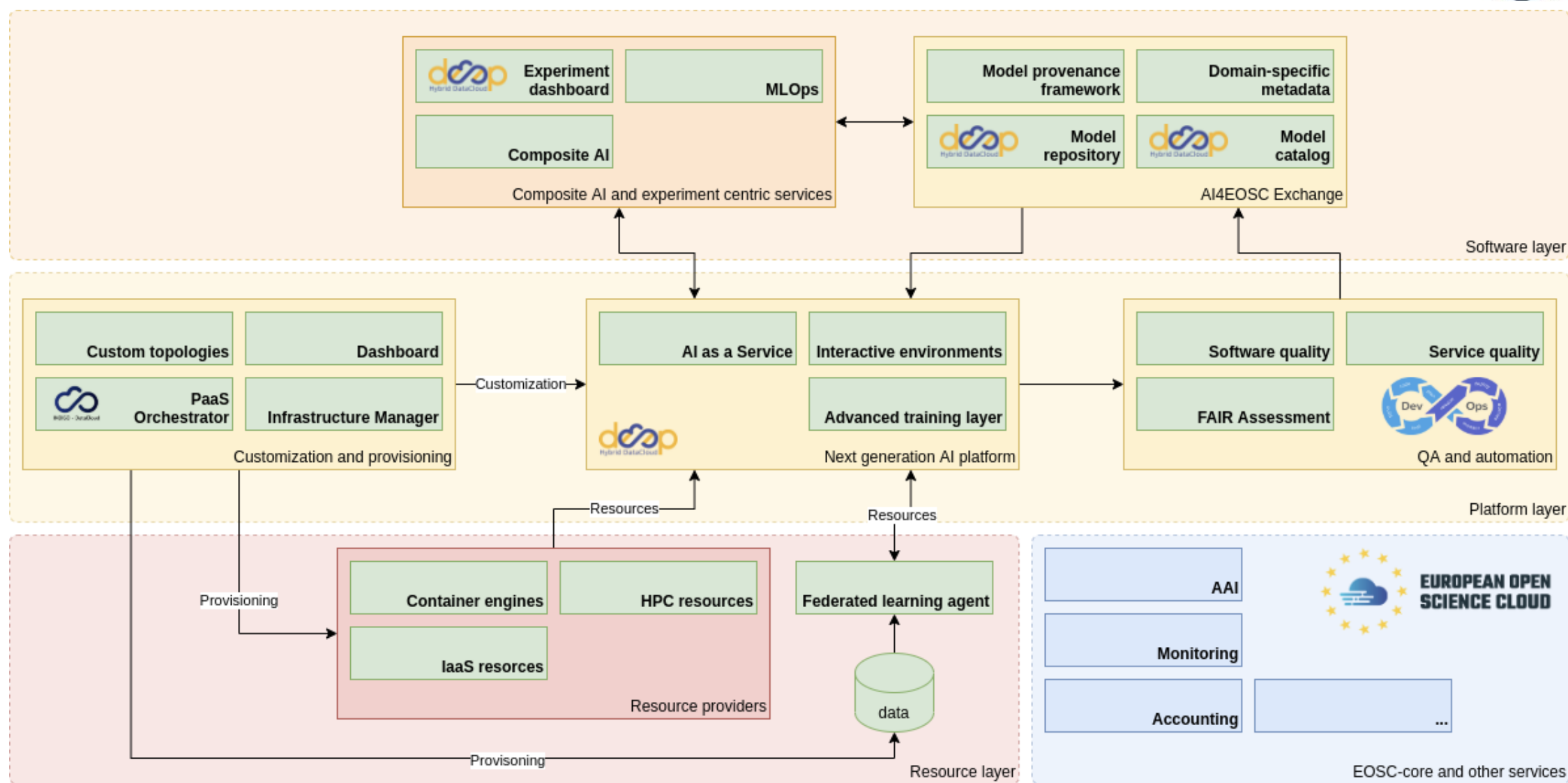
List of participants

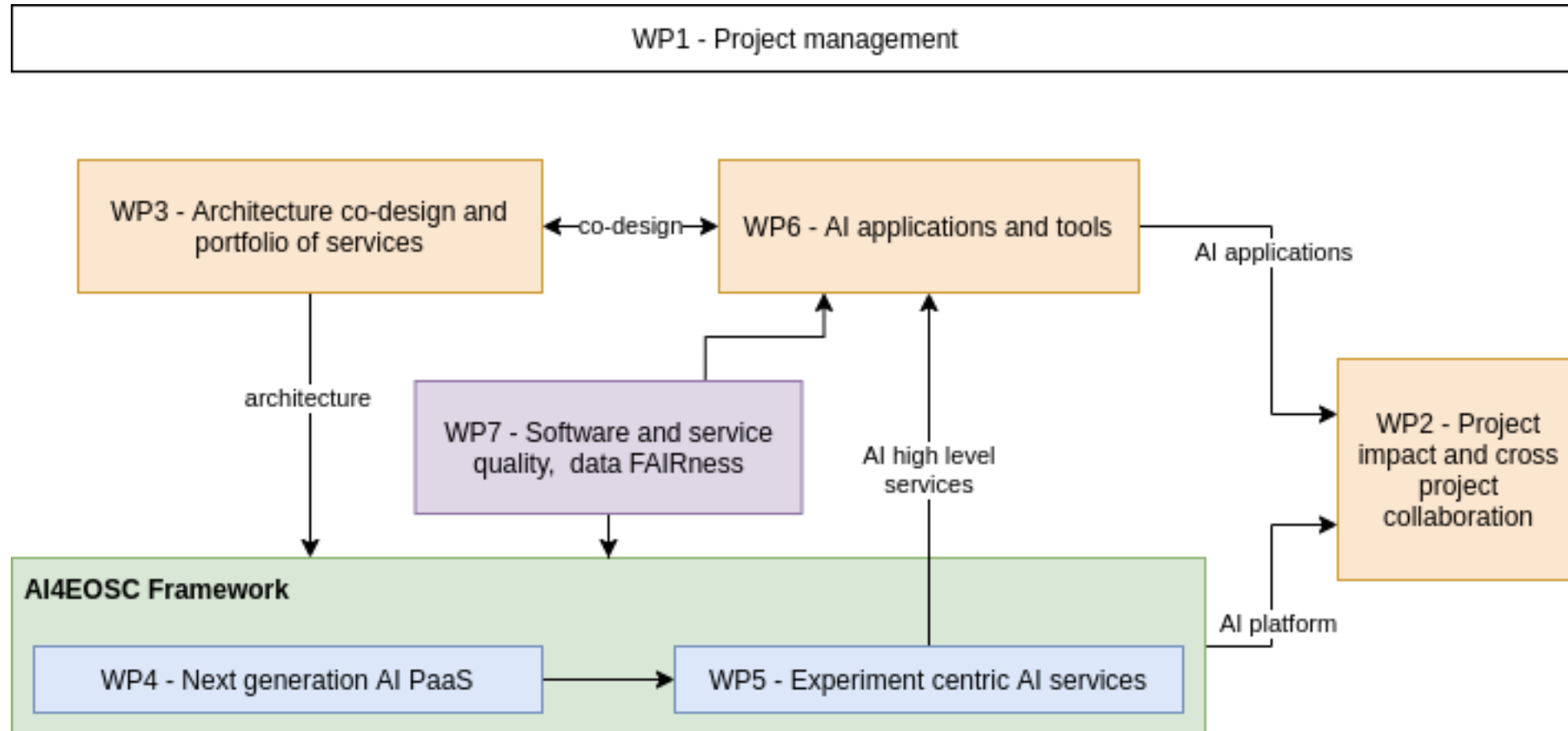
Participant No	Participant organisation name	Participant short name	Country
1 (coord.)	Agencia Estatal Consejo Superior de Investigaciones Científicas	CSIC	Spain
2	Karlsruher Institut für Technologie	KIT	Germany
3	Ústav informatiky, Slovenska akadémia vied	IISAS	Slovakia
4	Universitat Politècnica de València	UPV	Spain
5	Predictia Intelligent Data Solutions SL	PRED	Spain
6	Laboratorio de Instrumentação e Física Experimental de Partículas	LIP	Portugal
7	Istituto Nazionale di Fisica Nucleare	INFN	Italy
8	Instytut Chemii Bioorganicznej PA	PSNC	Poland
9	MicroStep-MIS, s.r.o.	MMIS	Slovakia
10	Wielkopolski Ośrodek Doradztwa Rolniczego w Poznaniu	WODR	Poland

- The vision of the AI4EOSC project is to increase the service offer in the EU landscape by expanding the European Open Science Cloud (EOSC) ecosystem to support the effective utilization of state-of-the-art AI techniques by the research community. In this regard, our project will provide highly innovative services built on top of existing EOSC services, thus allowing EU researchers to efficiently exploit large and distributed datasets, following a service-oriented approach over the EOSC continuum.
- The AI4EOSC project bases its activities on the technological framework delivered by the DEEP-Hybrid-DataCloud H2020 project.
- AI4EOSC will enhance this platform, delivering new high-level services and functionalities, targeting direct exploitation by scientific teams, allowing them to reduce the time to results and increase productivity by building better analytics tools, products, and services leveraging artificial intelligence, machine learning, and deep learning (AI/ML/DL), with focus on advanced features like federated learning, split learning or distributed training.

- **Objective 1:** Provide machine learning practitioners with feature rich services to build and deploy customizable machine learning, deep learning and artificial intelligence applications following a platform and serverless approach with horizontal scalability over the EOSC continuum
- **Objective 2:** Enhance existing cloud services to support machine learning and deep learning on distributed datasets, with a particular focus on federated learning.
- **Objective 3:** Deliver methods to build and compose machine learning and deep learning tools, making possible the development of more complex data-driven composite AI applications.
- **Objective 4:** Foster a machine learning and deep learning exchange in the context of the European Open Science Cloud, enhancing and increasing the application offer currently available in the DEEP Open Catalogue.
- **Objective 5:** Extend the service offer and the capabilities being offered through the EOSC portal, coordinating with the operational and management activities carried out by existing and future initiatives, creating and establishing cooperation synergies whenever possible.

AI4EOSC





AI4EOSC

Table 8: List of Work Packages

Work Package No.	Work Package Title	Lead Participant No.	Lead Participant Short Name	Person-Months	Start Month	End month
WP1	Project management	1	CSIC	61	1	36
WP2	Project impact and cross project collaboration	7	PSNC	59	1	36
WP3	Architecture co-design and portfolio of services	2	UPV	69	1	36
WP4	Next generation AI PaaS	6	INFN	200	4	36
WP5	Experiment centric AI services	1	CSIC	132	4	36
WP6	AI applications and tools	5	KIT	169	1	36
WP7	Software and service quality, data FAIRness	4	LIP	70	1	36
Total PMs				760		

Table 12 (3.1f): Summary of staff effort.

Part. No.	Participant short name	WP1	WP2	WP3	WP4	WP5	WP6	WP7	
1	CSIC	27	8	12	36	38	14	26	161
2	KIT	5	4	10	14	6	51	0	90
3	IISAS	3	2	8	48	30	8	15	114
4	UPV	6	2	13	34	18	0	10	83
5	PRED	5	9	4	0	6	22	0	46
6	LIP	3	2	4	8	18	0	19	54
7	INFN	5	2	6	60	16	0	0	89
8	PSNC	5	20	8	0	0	14	0	47
9	MMIS	1	5	2	0	0	33	0	41
10	WODR	1	5	2	0	0	27	0	35
TOTAL									760

Project Budget: 5M€

INFN Budget: 500k€

Sedi coinvolte: Bari, CNAF

PROJECT STARTS: 01.09.2022

DURATION: 36 Months

AI4EOSC: INFN Role



- Elicit the functional and non-functional requirements from the use cases that will guide the definition of the AI platform architecture.
- Establishment of the interaction loops between use cases and technical work packages (WP4, WP5 and WP6) to foster co-design the AI platform architecture To provision a distributed computing platform for secure and efficient training and inference of AI models over the EOSC continuum.
- To allow the dynamic deployment of AI services and applications across multiple back-ends, providing access to specialised acceleration hardware such as GPUs.
- To allow user communities to bring their own computing and storage resources and add them to the DEEP PaaS platform for training and inference.
- To allow the deployment of the whole, or parts of, DEEP stack on-premises, on public clouds or on federated resources.
- To support advanced learning scenarios: federated, split, incremental and distributed learning, to cooperatively train models across multiple datasets without requiring a shared data repository.
- To provide the computational support to easily expose trained models from the AI4EOSC Exchange as scalable inference endpoints in both on-premises and public Clouds.
- To deliver interactive development environments, allowing users to interact with the platform and get access to the underlying resources in an easy way.
- Composite AI through container and serverless function orchestration
- Creating a marketplace of AI services, applications, models and tools for the EOSC

Attività in fase di valutazione/negoziiazione

- Delibera del Comitato interministeriale per la programmazione economica (CIPE) n. 15 del 28 febbraio 2018 a valere sul Fondo per lo Sviluppo e la Coesione (FSC) 2014-2020, con una dotazione finanziaria complessiva pari a 200 milioni di euro, si inserisce nel processo unitario di programmazione strategica e operativa per il periodo 2014-2020, in coerenza con la Strategia Nazionale di Specializzazione Intelligente (SNSI) - area di specializzazione “Salute, alimentazione e qualità della vista”, il Programma Nazionale per la Ricerca (PNR) e il Piano Attuativo Salute della SNSI.
- È articolato in cinque Traiettorie di sviluppo cui corrispondono specifici Avvisi pubblici nell’ambito delle relative Linee di azione:
 - Traiettorie 1 “Active & Healthy Ageing - Tecnologie per l’invecchiamento attivo e l’assistenza domiciliare”;
 - **Traiettorie 2 “eHealth, diagnostica avanzata, medical device e mini invasività”;**
 - **Traiettorie 3 “Medicina rigenerativa, predittiva e personalizzata”;**
 - **Traiettorie 4 “Biotecnologia, bioinformatica e sviluppo farmaceutico”;**
 - Traiettorie 5 “Nutraceutica, nutrigenomica e alimenti funzionali”.

Accordo di ricerca con il Sant'Orsola di Bologna



- Studio, test e implementazione, a partire dalle soluzioni già in uso per la Piattaforma di Genomica Computazionale di IRCCS AOU e dalla piattaforma di analisi dati genomici sviluppata da INFN nell'ambito dell'Associazione del Ministero della Salute Alleanza Contro il Cancro, di un sistema di database genomico utile all'applicazione in progetti genomici large-scale;
- Realizzazione di una applicazione per il trasferimento dei dati genomici secondo standard di privacy by-design e by-default, automazione e integrità;
- Studio e applicazione di soluzioni basate su GPU (Graphical Processing Units) in metodi di analisi genomica, con l'obiettivo di migliorare la resa computazionale e la scalabilità;
- Studio e progettazione di piattaforme Cloud federate ed integrate per la gestione e analisi di dati omici;
- Adattamento di pipeline di calcolo ad architetture di tipo Cloud e Data Lake basate su microservizi;
- Esplorazione di scenari di integrazione con progetti analoghi condotti a livello nazionale ed europeo (p.es. 1Million Genomes, Health Big Data, ...) e con iniziative nell'ambito del progetto nazionale di ripresa e resilienza (PNRR), nonché di tecniche avanzate per la garanzia della privacy nell'analisi dati (privacy-preserving federated analytics, federated learning, multiparty homomorphic encryption, blockchain);
- Valutazione dei potenziali fabbisogni architetturali necessari all'integrazione tra dati omici e dati utili all'arricchimento delle analisi sopra descritte provenienti da altre fonti sanitarie quali sistema di Cartella Clinica Elettronica, LIS e RIS.

Conclusioni e considerazioni



- Come già avvenuto in passato questi progetti finanziati permettono di lavorare su argomenti/strumenti che poi sono utilizzati in produzione
 - L'orchestratore usato in INFN-Cloud è frutto di questo ciclo di sviluppo e ancora nei prossimi progetti sarà un componente centrale nell'attività prevista
- Prevediamo che le attività di **ML_INF**N possano beneficiare pesantemente
- Ci sono in corso PON/CIR che provvedono l'infrastruttura di base e il personale per renderla operativa, su cui questi progetti potranno insistere.
 - Vista la diversità delle risorse di calcolo (GPU, Fat Node, etc) le attività previste in questi progetti sono molto utili nell'ambito dell'**AI**
- Molte delle tecnologie pianificate in questi progetti potrebbero finire come «input» nei progetti del PNRR (CN-HPC, RI, PE) dove invece, dobbiamo integrare servizi già a TRL8 o maggiori
 - La capacità di applicare queste tecnologie a use case diversi da quelli core-INFN sarà un valore aggiunto per supportare le comunità (extra-INFN) dei progetti PNRR
 - Anche nei proposal PNRR, si faranno investimenti hw e sw a supporto della applicazioni che fanno uso di algoritmi di **AI**