Stato e attività di R&D nei progetti finanziati

Giacinto Donvito (donvito@infn.it)

Agenda

- Progetti approvati in partenza:
  - InterTwin
  - AI4INFN
  - EuroScienceGateway
- PON / CIR:
  - IBISCO
  - LifewatchPLUS
  - CNRBiomics
- Conclusioni e considerazioni

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

Target TRL: 6

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.

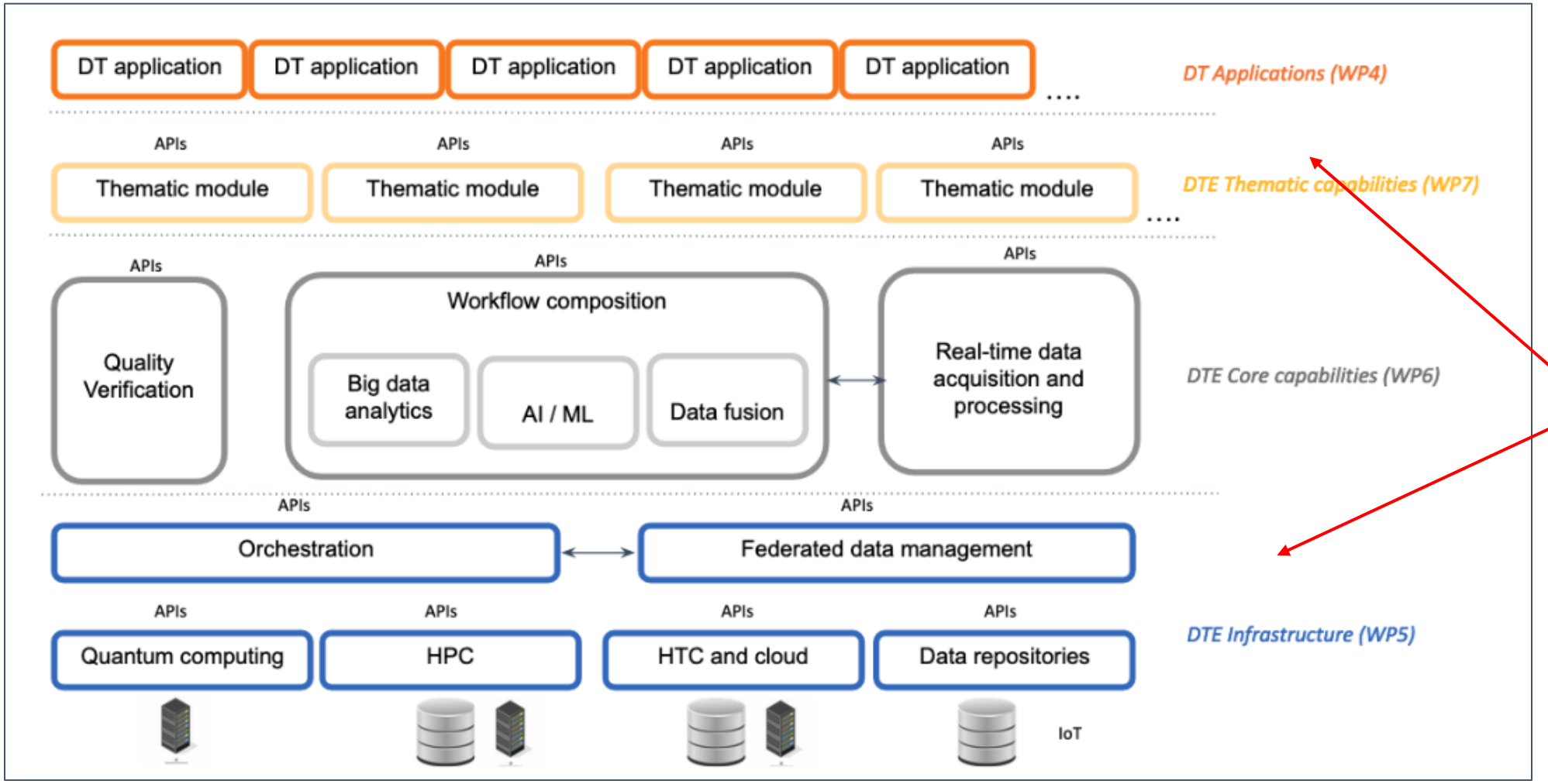
Objective 3. Extend the technical capabilities of the European Open Science Cloud with modelling & simulation tools integrated with its compute platform

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)**

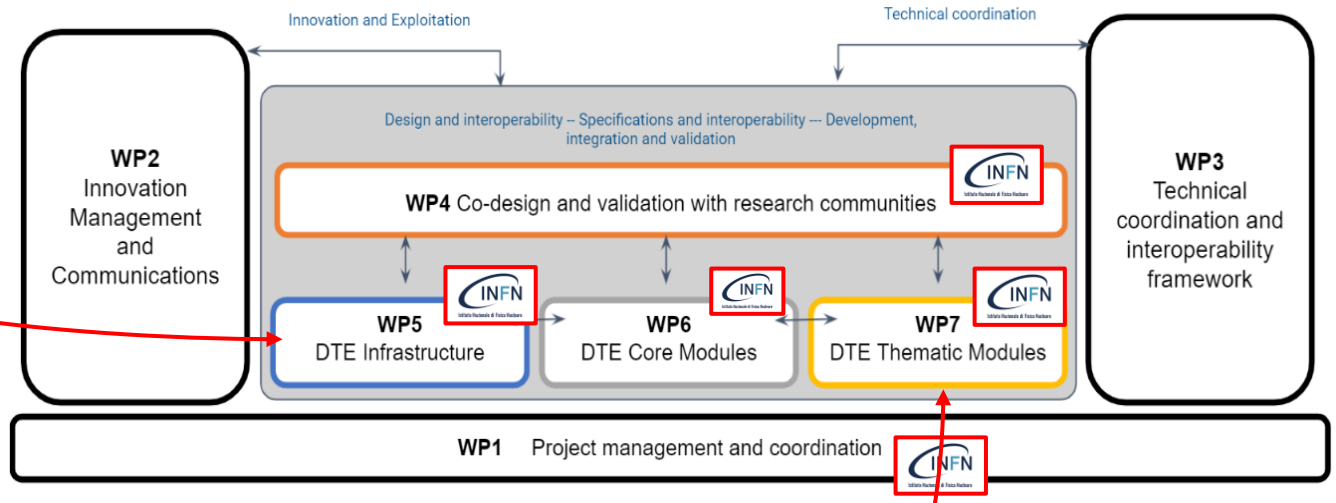
The DTE infrastructure provides the federated data and compute resources involved in modelling and simulation

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

Tech activities planned

Target TRL: 6

Federated compute infrastructure and data infrastructure:

- Computational task offloading from to HPC, Initial integration with the Federated Data Infrastructure
- PoC of manual file transfer between a data lake and one HPC facility
- Initial support to batch and interactive analysis and integration of federated AAI
- Manual file mirroring between data lake and storage at HTC, HPC and Cloud facilities
- Automated compute workflow integration (access policies and AAI) and data caching
- Solution for full-chain automated data management integrated with higher-level management and/or thematic services. On-demand file transfer

PaaS Orchestrator:

- Initial integration of orchestration and data management for AI-based task placement
- Algorithms for task placement using data locality information
- Integration of HPC nodes using automated dynamic data caching, use of network performance and full integration with data, HTC, HPC and Cloud

Big Data Analytics and workflow management

- Support of the application requirements for on-demand cloud and HPC resources
- Loosely integrated workflow engines and initial common workflow description language and serverless framework
- On-demand deployment of elastic and distributed Big Data processing kernels over clouds matching the requirements
- Hybrid workflows, AI/ML processes and enhanced data models. Serverless computing for real-time acquisition processing. Integration with Orchestrator for services/application deployment
- Integration (self-containerisation and same job execution specification) of parallel on-demand workloads over different backends
- Integration with thematic workflow engines. Workflow reproducibility and multi-level provenance management. Enhanced support for common real-time acquisition and processing. Full integration of the Orchestrator features

AI4EOSC

Giacinto Donvito

AI4EOSC

- 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 Fisica Experimental de Particulas	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

AI4EOSC

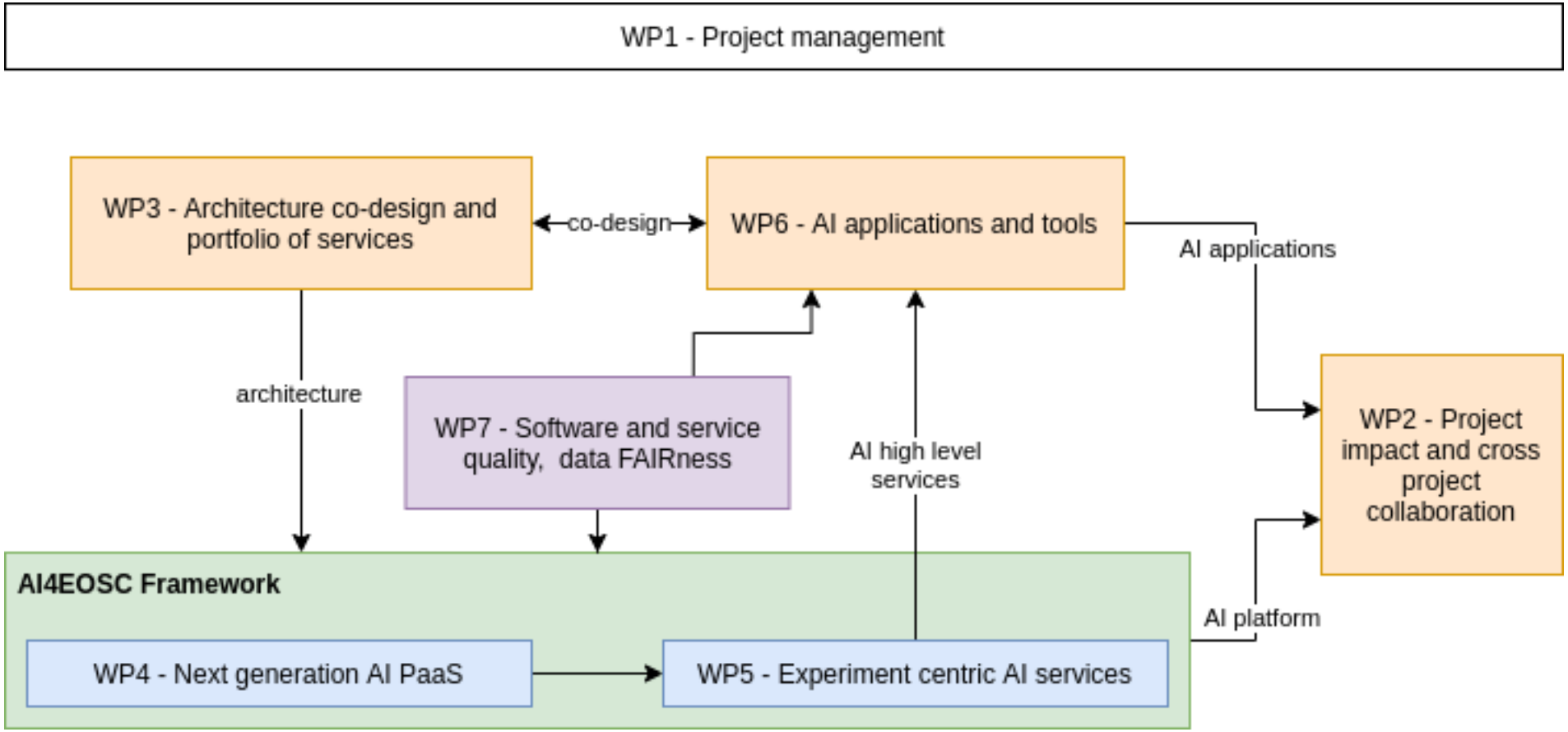
- The vision of the AI4EOSC project is to increase the service offer in the EU landscape by expanding the European Open Science Cloud (EOSC) ecosystem to support the effective utilization of state-of-the-art AI techniques by the research community. In this regard, our project will provide highly innovative 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.

AI4EOSC

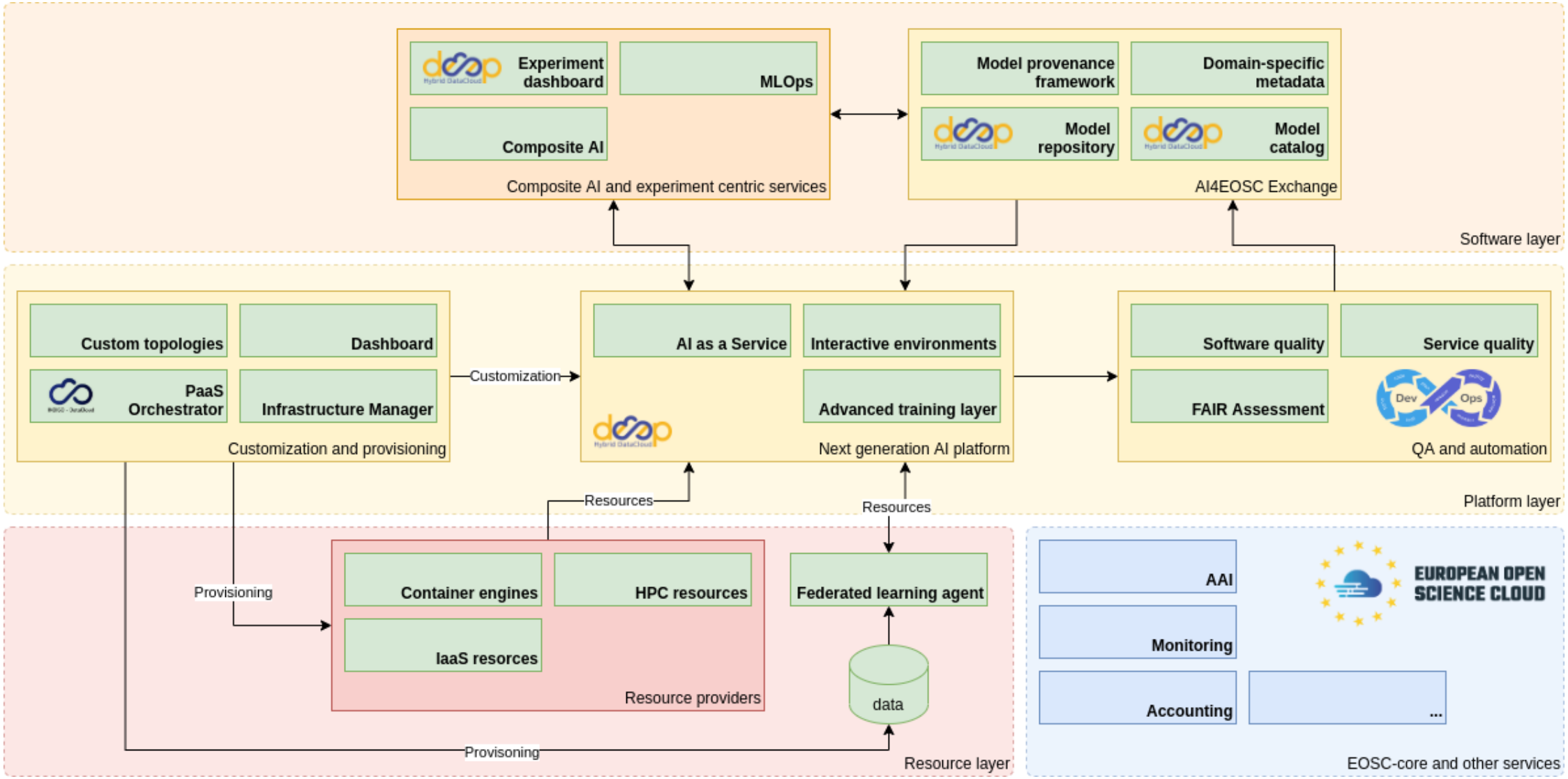
Target TRL: 8

- **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


WP1 - Project management
WP3 - Architecture co-design and portfolio of services
co-design
WP6 - AI applications and tools
AI applications
architecture
WP7 - Software and service quality, data FAIRness
AI high level services
WP2 - Project impact and cross project collaboration
AI4EOSC Framework
AI platform
WP4 - Next generation AI PaaS
WP5 - Experiment centric AI services

AI4EOSC


Experiment dashboard
MLOps
Composite AI
Composite AI and experiment centric services
Model provenance framework
Domain-specific metadata
Model repository
Model catalog
AI4EOSC Exchange
Software layer
Custom topologies
Dashboard
PaaS Orchestrator
Infrastructure Manager
Customization and provisioning
Customization
AI as a Service
Interactive environments
Advanced training layer
Next generation AI platform
Software quality
Service quality
FAIR Assessment
Dev
Ops
QA and automation
Platform layer
Resources
Resources
Provisioning
Container engines
HPC resources
IaaS resorces
Resource providers
Federated learning agent
data
Provisoning
Resource layer
AAI
Monitoring
Accounting
...
EUROPEAN OPEN SCIENCE CLOUD
EOSC-core and other services

Tech activities

Target TRL: 8

- To implement (by means of INDIGO PaaS Orchestrator) AI dedicated sw stack over both IaaS and docker orchestration facilities (Kubernetes) both public and private ones
- To support within INDIGO PaaS Orchestrator more data management solutions to serve the project use cases.
  - This will improve the flexibility and generality of the data management capabilities of INDIGO PaaS Orchestrator
- To implement an automatic deployment of the INDIGO PaaS core services.
  - This will help the user communities to deploy (and maintain) their specific instance of PaaS Orchestrator.
- To implement a simple solution for “automatic” addition of Computing and Storage Resources into a PaaS Orchestrator instance
- This task is responsible for providing the needed improvements to deliver an AI platform providing distributed, iterative, incremental, split and federated learning. The outcome of this task will be a set of software components to build a robust technological platform with a strong algorithmic focus.
- This task will also implement the re-training of the models based on at least two different logic: event based and time-based. In the case of event-based re-training this could be also triggered by configuring a threshold for running the re-training. In the case of the time-based re-training it will be possible to reexecute the workflow depending on the use case requirements.

Tech activities

Target TRL: 8

- To build an event-driven serverless inference framework based on the Functions as a Service (FaaS) model taking into account the requirements from WP6. The architecture will deal with the delivery of auto-provisioning and auto-scaling of FaaS frameworks (e.g. OSCAR, Knative) on top of provisioned Container Orchestration Platforms (e.g. Kubernetes), providing an elastic multi-tenant computing framework able to react to variable workloads and demands. These features will be built on top of existing serverless frameworks that will be extended to cope with the compute-intensive requirements of deep learning applications and support workflows of functions to facilitate the adoption of composite AI.
- This task will deliver an authenticated and interactive development and prototyping environments for application developers with access to e-Infrastructure resources. These environments will be based on state-of-the-art development platforms at tools (i.e. JupyterLab) and are focused on fast prototyping and development of applications, with limited access to resources. The aim of this task is to provide a familiar entry path for developers, before transitioning to the training/evaluation/testing phases done at scale. In order to avoid data loss (as these environments will be limited in time) we will offer syncing mechanisms so that the developer work can be resumed after the session ends. To this aim, we will exploit the resources offered by the platform.
- 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.

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

EuroScienceGateway

Giacinto Donvito

EuroScienceGateway

- Leveraging the European compute infrastructures for data-intensive research guided by FAIR principles

- 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.

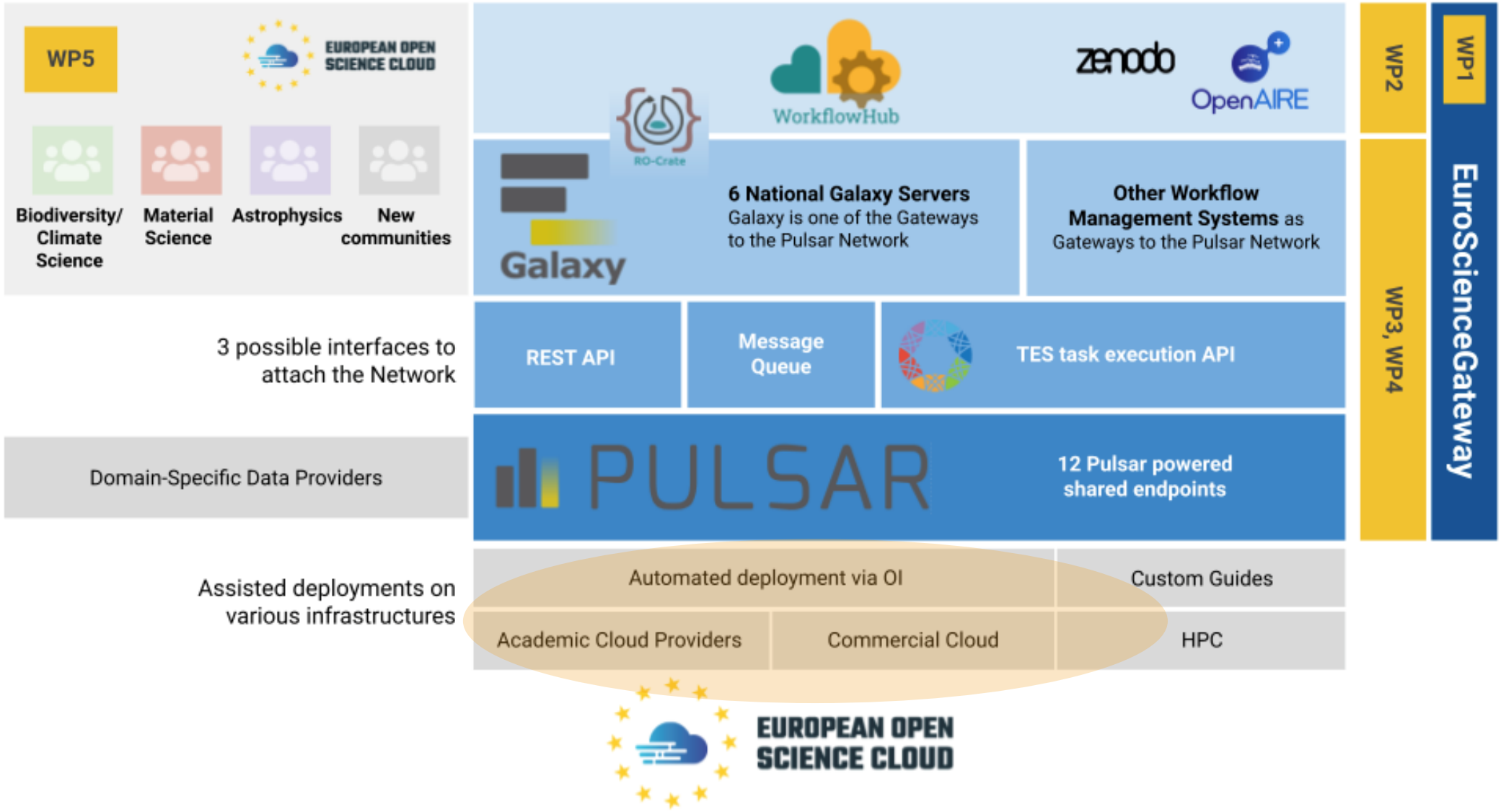
EuroScienceGateway

Project Budget: 4.5M€
INFN Budget: 125k€
Sedi coinvolte: Bari
PROJECT STARTS: 01.09.2022
DURATION: 36 Months

List of participants

Participant No. *	Participant organisation name	Country
1 (Coordinator)	Albert-Ludwigs-Universitaet Freiburg (ALU-FR)	DE
2	VIB VZW (VIB)	BE
3	Swiss Federal Polytechnical Institute (EPFL)	CH
4	Czech Education and Scientific NETwork (CESNET)	CZ
5	Barcelona Supercomputing Center (BSC)	ES
6	University of Barcelona (UB)	ES
7	Centre national de la recherche scientifique (CNRS)	FR
8	University of Paris (UP)	FR
9	Consiglio Nazionale Delle Ricerche (CNR)	IT
10	Istituto Nazionale Di Fisica Nucleare (INFN)	IT
11	University of Florence (UNIFI)	IT
12	European Grid Infrastructure (EGI)	NL
13	Universitetet i Oslo (UiO)	NO
14	AGH University of Science and Technology (AGH / AGH-UST)	PL
15	Ustav Informatiky, Slovenska Akademia Vied (IISAS)	SK
16	Scientific and Technological Research Council of Turkey (TUBITAK)	TR
17	United Kingdom Research and Innovation (UKRI), Science and Technology Facilities Council	UK
18	University of Manchester (UNIMAN)	UK

EuroScienceGateway



EuroScienceGateway: obiettivi

Target TRL: >=8

- **Objective 1:** Accessible e-Infrastructure resources for European scientists to enable pioneering data-driven research across scientific domains.
- **Objective 2**: Support the varieties of analysis types and diverse usage patterns through efficient and smart job distribution to appropriate and sustainable infrastructures.
- **Objective 3**: The application of FAIR principles to workflows and adoption of FAIR Digital Objects to stimulate reusable and reproducible research and enable the EOSC Interoperability Framework.
- **Objective 4:** Adoption of the EuroScienceGateway by researchers in diverse scientific disciplines.

EuroScienceGateway

- Task 3.1 Develop and maintain an Open Infrastructure based deployment model for Pulsar endpoints
  - We will further develop the OI to include the most used commercial clouds like Amazon Web Service (AWS), Microsoft Azure and Google Cloud Platform, as well as container orchestrators like Kubernetes. We will include EOSC-compliant AAI to facilitate integration with other EOSC services. This will enable the possibility to seamlessly deploy new Pulsar nodes directly from the Galaxy user interface (T4.1) for a wide range of Cloud solutions
- Integrating CVMFS, EGI Cloud infrastructures, into Pulsar network
- Develop an integration point, where a user can add compute resources to the personal Galaxy user account, e.g. AWS resources, or a Pulsar instance as an interface for local computing resources

- P.S.
  - Ottima sinergia con almeno 2 collaborazioni (JRU) a livello Italiano: Elixir // Lifewatch
  - Che si sono esplicitati anche in 2 PON infrastrutturali e 2 CIR (Personale)

PON:

- ➔Infrastruttura
- ➔Personale

PON: infrastrutture

- Le gare sono quasi finite tutte
- Abbiamo ottenuto un'ulteriore proroga:
  - Fino a Giugno 23
- Sono ancora in corso:
  - UNIBA:
    - Tape Library
    - Trigeneratore
    - Cablaggio LAN
  - INFN:
    - Firewall evoluto
    - Sala realtà aumentata
- La grande parte delle risorse pledged possono essere usate già dagli utenti

PON: Personale

- Ci sono ancora diverse posizioni da coprire,
- IBISCO:
  - Supporto allo sviluppo di applicazioni scientifiche che fanno uso dell'Infrastruttura IBISCO
  - Supporto alla governance di progetto
  - Gestione dei servizi dell'Infrastruttura IBISCO e sviluppo di applicazioni scientifiche
  - Integrazione di applicazioni e sviluppo di servizi multidisciplinari sulla infrastruttura IBISCO, formazione degli utenti e supporto alla preparazione e realizzazione di progetti di ricerca regionali, nazionali ed internazionali
  - Sviluppo di soluzioni di livello PaaS per l'esecuzione di applicazioni scientifiche su una infrastruttura di cloud distribuita
  - Studio e implementazione di soluzioni di federazione basate sulle tecnologie cloud per interconnettere le sedi del progetto IBiSCo
  - Modellizzazione e benchmarking di un cluster Infiniband con GPU
  - Utilizzo dello storage per applicazioni nel campo del patrimonio culturale
  - Sviluppo di un sistema di collaborazione per il calcolo scientifico
  - Realizzazione di una piattaforma Cloud su OpenStack
  - Studio e sviluppo di applicazioni di intelligenza artificiale per l'analisi di Big Data in ambiente di cloud computing

PON: Personale

- IBISCO:
  - Calcolo ad alte prestazioni su piattaforma ibrida CPU/GPU per attività di Machine Learning
  - Efficientamento energetico e monitoraggio dei data center IBISCO
  - Studio di sistemi distribuiti di archiviazione dati in ambito scientifico
  - Supporto alla creazione e gestione di archivi digitali ad accesso aperto per i prodotti della ricerca insistenti sull'Infrastruttura IBISCO
  - Sviluppo di sistemi per il calcolo scientifico e per l'analisi dati nell'infrastruttura IBISCO
  - Sviluppo e gestione di servizi per applicazioni scientifiche insistenti sull'infrastruttura IBISCO
  - Sviluppo e integrazione, nell'infrastruttura Ibisco, di soluzioni per la realizzazione di federazioni di dati, basate su software open source.
  - Modellizzazione e calcolo ad alte prestazioni su piattaforma ibrida CPU/GPU
  - Gestione e utilizzo di sistemi HPC per l'elaborazione di dati SAR satellitari
  - Gestione e analisi dati su server di calcolo ad alte prestazioni
  - Supporto allo sviluppo di applicativi scientifici per l'astronomia Cherenkov che fanno uso dell'Infrastruttura IBISCO
  - Studio e sviluppo e ottimizzazione di complessi workflow di analisi di dati su infrastrutture di cloud computing distribuite.

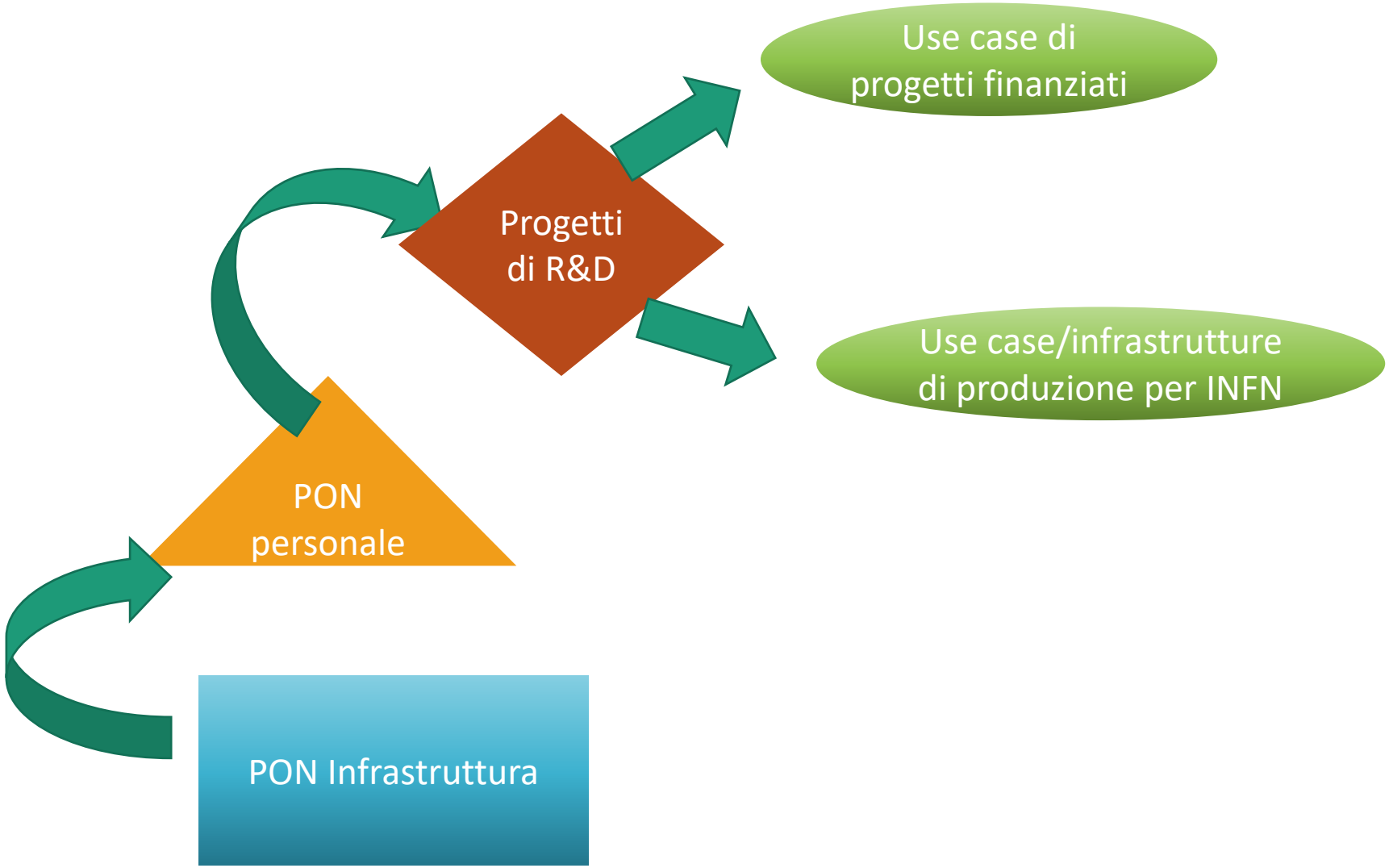
PON: Personale

- 2 in corso di assunzione, 3 ancora da coprire
- LifewatchPLUS:
  - Installazione, configurazione e mantenimento di una infrastruttura cloud basata su OpenStack
  - Installazione, configurazione e mantenimento di una infrastruttura HPC per la Big Data analytics.
  - Studio e implementazione di tecnologie di cloud computing and storage per la federazione di data center distribuiti per l'esecuzione di applicazioni di Big Data Analytics
  - Sviluppo di un Virtual Research Environment per l'esecuzione delle applicazioni scientifiche in una infrastruttura di cloud distribuita
  - Sviluppo di strumenti di data cube per l'organizzazione dei dati di Osservazione della terra per l'estrazione di variabili utili al monitoraggio della biodiversità biodiversità e dello stato di conservazione degli ecosistemi

PON: Personale

- Ancora 3 posizioni da coprire
- CNRBiomics:
  - Implementazione di una piattaforma di calcolo per la Bioinformatica - Installazione, configurazione e mantenimento di una infrastruttura cloud basata su OpenStack.
  - Implementazione di una piattaforma di calcolo per la Bioinformatica - Installazione, configurazione e mantenimento di una infrastruttura HPC per la Big Data analytics
  - Implementazione di una piattaforma di calcolo per la Bioinformatica - Installazione, configurazione e mantenimento di una infrastruttura cloud basata su OpenStack - Installazione, configurazione e mantenimento di soluzioni di gestione di grandi moli di dati per le applicazioni scientifiche
  - Implementazione di una piattaforma di calcolo per la Bioinformatica - “Aspetti legali e normativi conseguenti la gestione e l'analisi dei dati sensibili in progetti di ricerca nazionali, internazionali ed Europei”.

Sinergie:


Use case di progetti finanziati
Progetti di R&D
Use case/infrastrutture di produzione per INFN
PON personale
PON Infrastruttura

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 in questi progetti è un componente centrale nell'attività prevista
- Prevediamo che le attività di ML_INFN 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
- 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 dei progetti PNRR