

Aggiornamento su attività del PO

Daniele Spiga

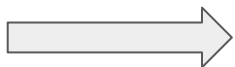
Program Officer

- Razionale:
 - I progetti esterni hanno durata in generale breve e natura molto variabile, con esigenze diverse; devono essere gestiti in modo dinamico
- Il PO Mantiene la “dashboard dei progetti”, assiste il coordinatore e il PMB nella definizione del programma di lavoro di DataCloud
- In contatto con il Working Group Progetti
 - DataCloud porta il contributo tecnico alla discussione nel WG Progetti e nel C3SN
 - Creazione di una Dashboard progetti in cui tracciare le caratteristiche e le scadenze

Impostazione del lavoro (da Febbraio)

Tre assets principali

- Raccolta milestones & deliverables dei progetti ICSC Spoke0 e Spoke8, TeRABIT, DARE, S.Orsola e Health Big Data
- Sviluppo delle singole milestones in sotto-attività (breakdown)
 - Identificazione di sinergie (dipendenze) tra le varie milestones
- In parallelo generazione dell'inventario di tutti (gli altri) progetti esterni sinergici a DataCloud
 - Principalmente progetti finanziati EU
 - In collaborazione con il WG Progetti



La finalità è definire milestones/deliverable interni a DataCloud

Stato raccolta Milestones

Tutte le **milestone** dei progetti identificati sono state **raccolte e organizzate in una unica tabella (la dashboard)**

- Intervallo 06.2024 - 10.2025
 - Le scadenze iniziali sono state perlopiù revisionate
- Dettagli discussi con i PI
- Situazione non statica soprattutto per progetti life science/medicali

Project	Deliverable/Milestone (full table here)	Original Date	Real Date	Type	Breakdown	Note
ICSC-S0	Activation of a PoC for resource federation	30 June 2024	30 June 2024	Integration/devel	BD1	
ICSC-S0	HAMMON D1.3 Implementation of the first PoC of the Cloud Platform	30 June 2024	TBD	Integration/devel	BD_H_1	
ICSC-S0	HAMMON D1.4 Implementation of the first integrated version of the Cloud Platform	31 October 2024	TBD	Integration/devel	BD_H_2	
ICSC-S0	End of project: Infrastructure in production	31 August 2025	31 August 2025	Integration/devel	BD2	
ICSC-S0	HAMMON D1.5 Implementation of the fully featured high-available Cloud Platform	31 August 2025	TBD	Integration/devel	BD_H_3	
TeRABIT	D4.3 - Report on the deployment of PoC applications over multiple HPC Bubbles	31 December 2023	30 June 2024	Integration/devel	Vedi BD1	Come BD1. Le Bubbles si emulano se non ci sono
TeRABIT	D4.4 - Report on the deployment of a PoC for dynamic caches via the PaaS Orchestrator	31 March 2024	30 June 2024	Integration/devel	BD_T_1	
TeRABIT	End of project: Infrastructure in production	30 June 2025	30 June 2025	Integration/devel	Vedi BD2	Come BD2
ICSC-S8	Migrare UISS da piattaforma bare metal a piattaforma cloud certificata e certificarlo come SaaS.	28 February 2024	30 August 2024	Integration/devel/data governance	BD_S8_1	
ICSC-S8	Lavorare sull'integrazione dei servizi cloud INFN e CINECA (in sinergia con Spoke0) a livello tecnico (IAM, Orchestratore, data management) e organizzativo (armonizzazione delle policy)	30 October 2024	30 October 2024	Integration/devel/data governance	BD_S8_2	
ICSC-S8	Avviare e concludere una PoC per replicare il lavoro svolto da A.Cavalli con Humanitas e Engineering sulla piattaforma ICSC. Avremo una prima call di avvio con il rettore di Humanitas il 5 marzo dove inizieremo a pianificare le attività. Obiettivo chiave della PoC sarà dimostrare che la piattaforma cloud ICSC e' in grado di gestire le molli di dati gestite attualmente da humanitas/engineering con performance equivalenti.	30 August 2025	30 August 2025	Integration/devel/data governance	BD_S8_3	per questo task dovremmo partire dalla PoC della riga precedente e dimostrare di essere in grado di gestire il trasferimento dei dati genomici dalla fonte (sequenziatori del progetto 5000 genomi di Aosta) alla piattaforma ICSC (CINECA+INFN). Il fatto che CINECA e INFN siano entrambi coinvolti e raggiungibili da un unico entry point DataCloud e' un requisito. L'idea e' di utilizzare RUCIO/FTS.
DARE	Integrare la piattaforma AlmaHealthDB in ICSC DataCloud (la relazione tra DARE e ICSC e' esplicitata nel progetto) e certificarla come SaaS.		30 October 2024	Integration/devel/data governance	BD_M_1	SaaS e' inteso come service level nello shared responsibility model standard. In pratica significa che noi ci occupiamo e siamo responsabili (o co-responsabili insieme a partner di UniBO e CINECA) del "contenuto del tenant". Più dettagli tecnici li avremo dopo il 23 aprile, data di una riunione tecnica Spoke8/DARE dove definiremo la proposta. Parlo di Spoke8/DARE perché le persone che lavorano su questo sono le stesse in entrambi i contesti
DARE	Integrare i progetti pilota DARE nella piattaforma ICSC e certificarli come SaaS - Alcuni progetti si riducono all'utilizzo di Ansys su cloud, altri necessitano di workflow manager come NextFlow (expertise di Gasparetto maturata con Sant'Orsola) Sara' probabilmente necessario integrare anche RedCap, gestito da un consorzio di cui fa parte anche il CINECA. Dovremo decidere come porci (entrare nel consorzio? integrare il software come "esterni"?)			Integration/devel/data governance	BD_M_2	
S. Orsola	Costruire una piattaforma di genomica computazionale cloud based e certificata come SaaS . La piattaforma integrera': Il software dell'ecosistema Elxir (Galaxy &C.)	Settembre '25	Settembre '25	Integration/devel/data governance	BD_M_3	
Health Big Data	IAM + PaaS Orchestrator + Rucio + Metadati Altri elementi che dovremo integrare in HBD sono RedCap (vedi DARE), e XNAT (usato anche in THE, su questo stiamo impostando una collaborazione con A. Retico) Aggiungo alcune informazioni ricevute da ACC (piattaforma principale in produzione su HBD) per l'evoluzione			Integration/devel/data governance	vedi BD_S8_2	

E relativo breakdown

Tutte le milestones sono state analizzate ed è stato realizzato il breakdown

- Include mapping sui Work Packages di DataCloud
- Necessaria review tecnica di CTO e PMB
- E' stato il punto di partenza per il **Workplan del PoC**

Project	Deliverable/Milestone (full table here)	Original Date	Real Date	Type	Breakdown
ICSC-S0	Activation of a PoC for resource federation	30 June 2024	30 June 2024	Integration/dev	BD1

Activation of a PoC for resource federation							
1- Federazione minimale di un provider Openstack							
2- Federazione a livello di applicazione via offloading							
Involved Areas	Involved Services/Components	Fine grained activities	Date	Priority	Owner	Contributors	NOTE
AAI		IAM PoC/Test Instance Deployment			Wp6		Questo sarà quello usato per tutti i servizi centrali e high level che faranno parte del PoC
		IAM(s) Operations			WP6	WP2/1	
Compute Federation	PaaS Orchestrator - 1. Openstack federation, - 2. Service level offloading	Deployment of multitenancy dashboard			WP5		
		Orchestrator configuration (i.e. federation registry) - Option 1			WP1	WP5	
		Federation validation (i.e. high level service deployment)			WP6		Per la federazione via Orchestratore vanno definiti i servizi high level da "usare" per il PoC
		High level service deployed by Orchestrator configured to offload toward external Provider (batch and other: vm or k8s)			WP6	WP5	Selezionando un servizio k8s puo' essere abilitato l'offloading nelle seguenti condizioni target. - Nodo (fat node istanzionato su un qualunque OpenStack) - batch - k8s
Data Federation	RUCIO	Actual Rucio(s) Deployment			WP6	WP1	Puo' essere un servizio dedicato sull'infrastruttura di WP6
	FTS	FTS Deployment (if needed, might reuse the existing one)			WP6	WP1	Si usa l'istanza INFN.
		Storage endpoint configuration at site (depend also on external providers specific setup)			WP6	WP1	Vanno definiti gli endpoint da federare. Quali e Quanti

Activation of a PoC for resource federation

- [Access and Identity Management](#)
- [Workplan](#)
- [Stima dei tempi:](#)
- [Compute Resource Federation](#)
- [Workplan:](#)
- [Federazione di istanze Cloud](#)
- [Service Level Offloading](#)
- [Stima dei tempi:](#)
- [Data Federation](#)
- [Workplan](#)
- [Stima dei tempi:](#)

Questo è un documento sintetico e operativo per la realizzazione del primo PoC per la federazione delle risorse ICSC. Tutte le scelte implementative qui presentate sono in linea con l'obiettivo e l'architettura finale ma, ovviamente, tengono conto dei tempi strettissimi del deliverable del 30 Giugno 2024. Pertanto sono state identificate molteplici configurazioni semplificate che saranno evolute in un secondo momento.

Le attività tecniche relative al cronoprogramma descritto in questo documento sono **quasi totalmente parallelizzabili**, ovvero sono attività atomiche e autoconsistenti e non hanno interdipendenze. Lato INFN si possono identificare quindi pool di persone diverse che devono supportare la fase realizzativa.

L'unica vera interdipendenza è la parte di Access and Identity Management, utilizzata sia nella federazione "compute" che in quella "data".

NOTA: In rosso sono indicate le azioni operative che prevediamo siano a carico di CINECA e che quindi vanno concordate con loro (oltre ovviamente a concordare con loro il piano globale). Queste saranno fuori del controllo diretto realizzativo dell'INFN.

Access and Identity Management

In questa prima fase, la proposta è di avere una istanza IAM unica (dedicata al PoC) che federi gli Identity Providers del CINECA e dell'INFN (INFN-AAI) e altri se necessario.

Workplan

- Deployment di uno IAM comprensivo del servizio VOMS-AA

Milestones/Deliverables interni

Primo tentativo di riorganizzazione in milestones interne

- Ancora non conclusivo

Alcune milestones hanno dipendenze che potrebbero essere bloccanti (vedi certificazioni)

- O comunque richiedere una rivisitazione

ID	DataCloud internal Deliverable/Milestone	Date	Datacloud WP (Responsible)	DataCloud WP (Collaborating)	Note
1	ICSC IAM Instance deployed	May 15th	WP6	WP5	serve asap per le altre Milestones. additional details here
2	PaaS Orchestrator / Dashboard ready to federate Galileo100 (CINECA)	June 15th	WP5	WP1	additional details here
3	interLink based offloading against Slurm CINECA and other backends (Bubbles emulation) ready	June 15th	WP6	WP1/WP5	Slurm puo' essere cineca. Per Bubble possiamo decidere altro backend (batch, fat node, k8s)
4	Data Federation ready. with at least 2 INFN endpoints federated via Rucio and FTS: Ready to federate CINECA	June 15th	WP6	WP1	additional details here
5	One cache technology selected for automated deployment via PaaS	June 15th	WP5	WP6	additional details here
6	Catchall Rucio Deployment Model definition		CTO	PMB	
7	Iam deployment model in the context of ICSC (i.e multi VO) and beyond ICSC	Jul. 2024	CTO	PMB	
8	Certificazione CT per supporto piattaforma UISS	Jul. 2024	WP7	WP1	
9	UISS Cloud Platform deployed	Aug. 2024	WP1	WP5	
10	Vulnerability assesment di UISS	Aug. 2024	WP7	WP4	
11	IAM ISO Certification /hardening				
12	PaaS ISO Certification/ hardening				
13	Rucio ISO Certification /hardening				
14	Deploy Spoke8 Compliant PoC (2FA, GDPR)	Oct. 2024	WP7	WP6	
15	K8s federation model design definition	Oct 2024	WP5	WP1	
16	PaaS Orchestrator fully enable k8s federation		WP5		
17	"Humanitas e Engineering test" with Rucio based datalake and dummy Data (involves ISO certd infra)	Aug 2025	WP6	WP7/WP1	
18	Full Compute Federation enabled	May 2025	WP1		
19	Full Data Federation Enabled	May 2025	WP1		
20	Multy/many IAM / IDP Deployed	May 2025			
21	Elixir ecosystem (Galaxy & Co) certified	Aug. 2025	WP7	WP4	
22	HAMMOND high level services configuration ready	TBD	WP5		
23	HAMMOND high level services integrated with PaaS	TBD	WP5		

Still Preliminary

Inventario progetti esterni (e sinergie)

- **Parte dell'attività PO meno matura**
 - Le classi di sinergie identificate sono una prima proposta
- Stessa logica di prima: creare una dashboard aggiornando dinamicamente le informazioni e i nuovi progetti
 - L'attuale schema considera i progetti attivi (input da Luciano)

			Sinergie del progetto verso DataCloud.										
Progetto	Sezioni coinvolte (con Budget)	PaaS Orchestrator	AuthN/Z - IAM	Data Management	Software Management	Modelli federazione/integrazione	Certificazione e ISO	Servizi di Alto Livello (sia PaaS che SaaS)	Soluzioni per Accounting	Policies	Hardware	PI	Note
interTwin	1. Bari 2. CNAF 3. Perugia 4. Pisa (sia infrastruttura che use case) 5. Torino (solo per la parte Use Case)	Adottato come middleware architetturali per il Digital Twin Engine: federazione Cloud Il progetto finanzia effort a INFN (Bari/CNAF) per 1. lo sviluppo della selezione intelligente del miglior provider 2. Contribuisce all'integrazione del meccanismo di offloading per i deployment di servizi basati su k8s				Il progetto finanzia effort a INFN (Perugia/Pisa/CNAF) per lo sviluppo di InterLink: un API Layer per abilitare l'offloading di pods (container) da un k8s verso un centro HPC / HTC ed eventualmente altro backend			Il progetto finanzia effort a INFN (CNAF) per lo sviluppo/integrazione di soluzioni per accounting di risorse. In particolare accounting di risorse fornite attraverso k8s			Spiga	
AI4EOSC	1. Bari 2. CNAF	Implementation and integration 1 Federation registry (replacement of CIP and CMDB) 2 Monitoring and metering 3 CPR						Implementation of automated deployment of Kubernetes				Donvito	
EOSC-beyond												Bozzi/Giacomini	
AARC												Giacomini	
Spectrum												Boccali	
ACME													

Considerazioni

Raccogliere tutte le informazioni ha richiesto un po più di tempo del previsto

- Anche andare nel merito delle singole attività non sempre immediato
- Cruciale l'interazione con chi ha lavorato alla preparazione dei progetti e ne segue direttamente le evoluzioni
 - In particolare per le interazioni/dipendenze con altri WP/partners

Approccio indubbiamente utile non solo per tenere traccia del lavoro ma anche per evidenziare eventuali criticità introdotte da interdipendenze

- Dipendenze da milestones complesse possono introdurre ritardi
 - Vedi processi di certificazioni di siti, middleware, processi etc

A valle di questa prima esperienza potrebbe essere utile adottare approccio analogo per le attività relative ai “progetti interni”