

Ipotesi di definizione del WG Infrastruttura

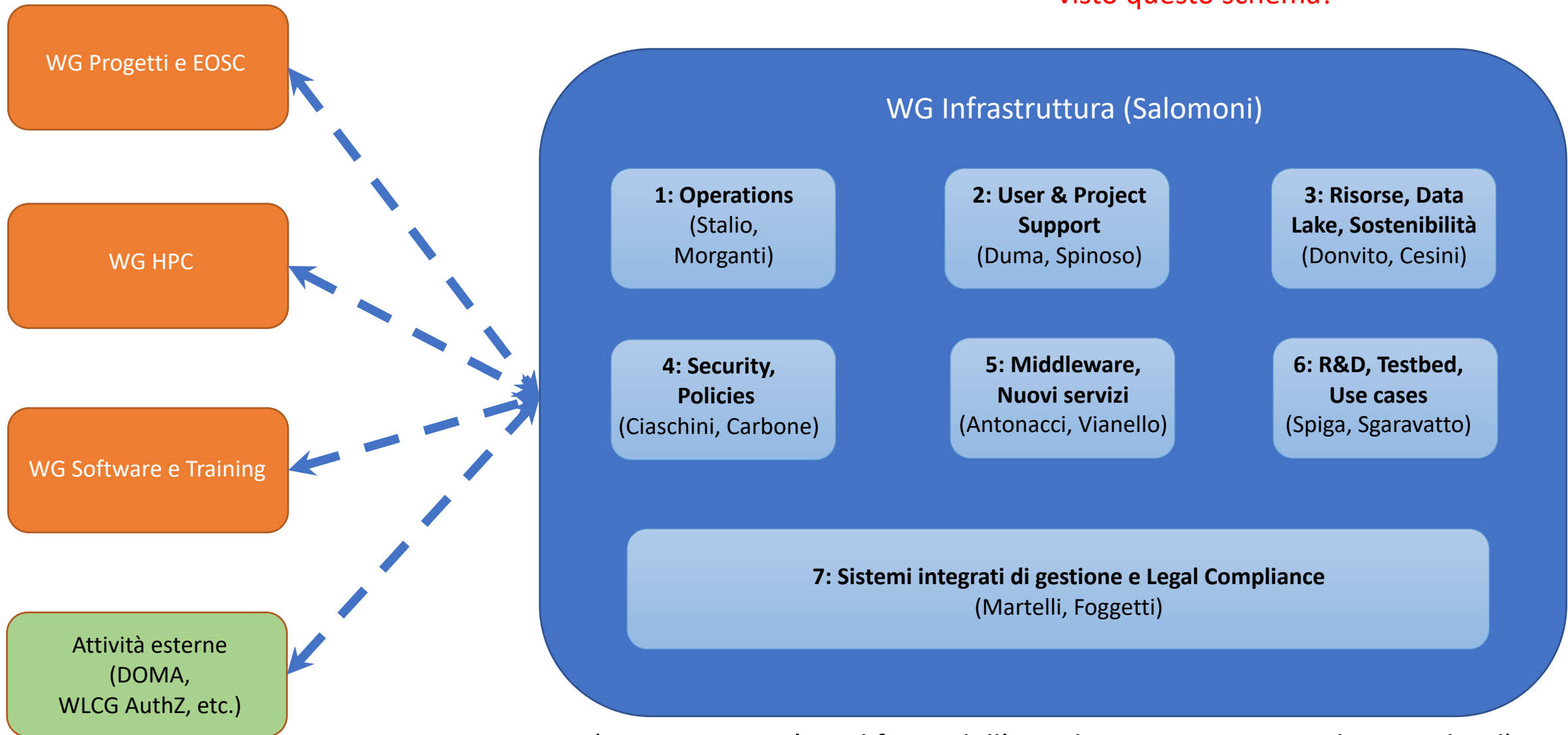
Davide Salomoni
Bologna, 1/4/2022

Mandato “broad” del WG Infrastruttura

- Il WG infrastruttura dovrebbe occuparsi di **diverse attività core del calcolo INFN**:
 - Sviluppo e gestione del modello architetturale di Data Lake
 - Sviluppo delle infrastrutture ISO-certified (che sicuramente avranno un “boost” anche a seguito di progetti PNRR)
 - Supporto alla gestione e all'operazione dei siti, agli esperimenti e agli utenti
 - Sviluppo nuovi servizi
- Un punto strutturale e fondamentale è **l'integrazione tra l'infrastruttura tradizionale «a Tier» tipica di WLCG e il modello «cloud-native»**.
 - **Integrazione**: di risorse, metodi, persone, soluzioni.

Possibile struttura del WG

Disclaimer: le persone indicate **non** sono state contattate e **non** hanno visto questo schema!



(versione un po' modificata dell'attuale struttura interna di INFN Cloud)

Alcuni “punti caldi” di questo WG

- Dove vanno le **risorse hardware** PNRR (Tier-2 e INFN Cloud) – vedi dopo
- Dove vanno le **persone** PNRR
 - **Centro Nazionale**: 29 TD per spoke infrastruttura + 2-3 TD per spoke omics data
 - **TeRABIT**: 20 TD
 - Arriveranno **progetti e risorse umane e/o hardware anche su altri bandi** (ad es. Partenariati estesi, piano complementare MUR-Min.Salute)
- Vanno **armonizzate**:
 - Le procedure di **reclutamento e la gestione del personale assunto**: bandi (centralizzati) e modalità di lavoro
 - Esempi: EGI, GARR
 - Le persone possono risiedere in una sede e lavorare “per il WG infrastruttura”. Deve essere prevista la possibilità di effettuare missioni (→ forse i costi di missione saranno più alti del solito). **Da chiarire bene il rapporto con i direttori di struttura**. Ad esempio:
 - Le persone assunte nella sede X lavorano “part-time” per il WG infrastruttura (es. 70/30)
 - Ogni Y persone che la sede X “ospita” e che lavorano full time per il WG infrastruttura, si dà alla sede X un A/R.
 - Dobbiamo trovare le strade migliori per il recruitment (esempio: <https://academicjobsonline.org>, selezioni ad hoc per il PNRR, etc.)
 - **Il personale assunto verrà associato ai WP della slide precedente.**
 - **L’uso delle infrastrutture** (incluse TeRABIT, Centro Nazionale, altre RI, etc.)

ICSC

Centro Nazionale HPC,
Big Data
e Quantum Computing

Cloud national infrastructure for supercomputing
Hub & Spoke organization:

10 vertical spokes for technology developments and software applications

400 M€ Total Budget

188 M€ Cloud Infrastructure

40 M€ Open Call

40 M€ Innovation & TT

42% Investment South Regions

34 MUR Universities
and Research institutions

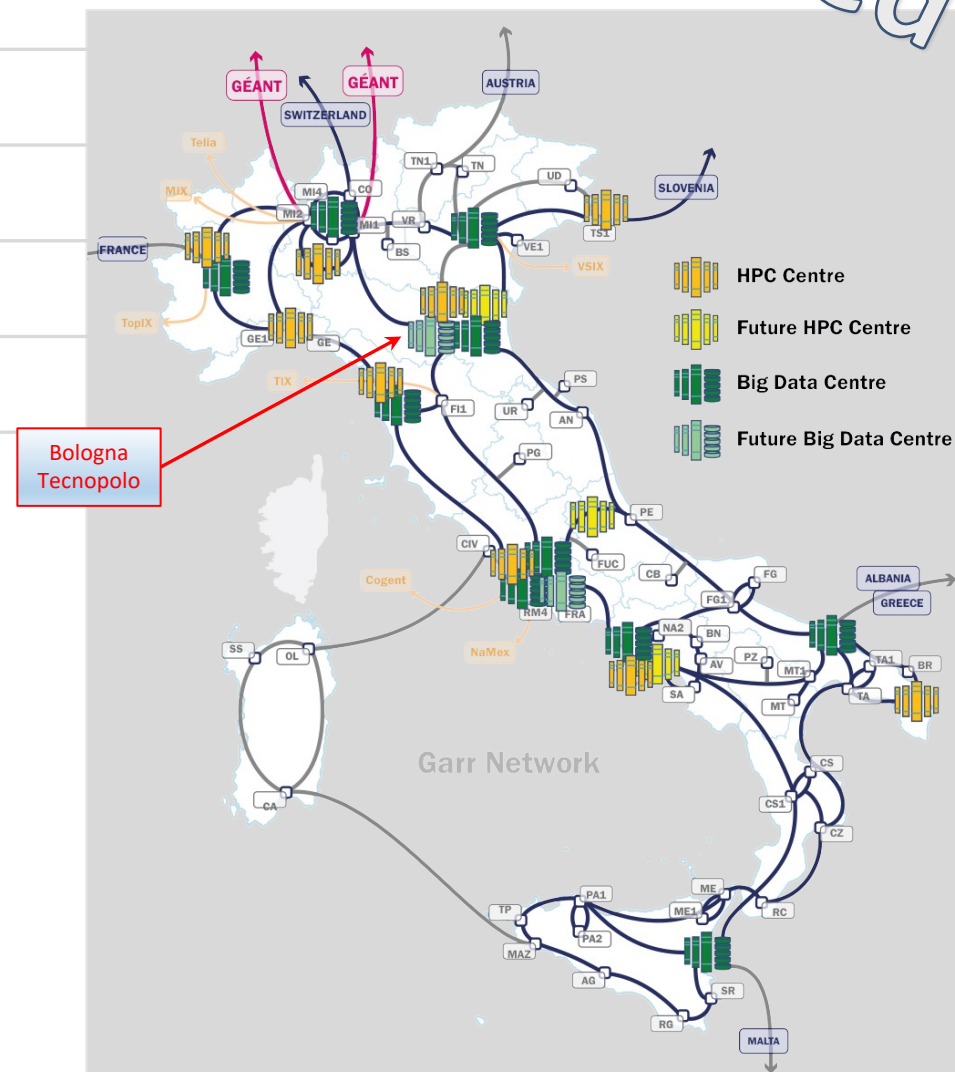
15 Private Companies

1575 Researchers and Engineers

250 New Temporary positions

250 New PhD

40 % Female



Proposed

TeRABIT (bando Infrastrutture di Ricerca)

- **Nel bando PNRR su “Infrastrutture di Ricerca”**, sono stati allocati 90M€ all’area “Data, Computing and Digital Research Infrastructures (DIGIT)”.
 - Il 27/2 è stata sottomessa su quest’area la proposta **TeRABIT**, con **capofila INFN**.
 - Partecipanti: **INFN** e **OGS** (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale). Partner “esecutori”:
 - GARR, per la realizzazione di connettività ad alta velocità al sud (associato a INFN).
 - CINECA, per l’upgrade di «PRACE-Italy» (associato a OGS).
 - Finanziamento complessivo richiesto circa 55M€, durata 30 mesi.
 - Per INFN, abbiamo richiesto un totale di circa 20M€.
- Per l’**INFN**, sui 20M€ previsti abbiamo richiesto circa **16.2M€ per hardware, interamente destinati a soluzioni Cloud** focalizzate in particolare sul concetto di “HPC Bubbles” (HPC distribuito e fruito in modalità cloud) e di distributed caches.
 - Circa il 20% al sud (3.3M€) e 80% al nord (12.9M€), a compensare la distribuzione 100% sud di Cloud nel Centro Nazionale.
 - Comprende la costituzione o l’espansione di cloud ISO-certified.
- Come **personale**, per l’**INFN** abbiamo richiesto circa **20 TD di durata biennale**, da dividere tra le sedi INFN che partecipano alla INFN Cloud.

Dove vanno le risorse

- Dal CN:
 - **30M€** per *infrastruttura* nei Tier-2 (1/3 per upgrade “non-ICT” dei siti)
 - **12M€ al sud**
 - **5M€ previsti per hardware Cloud services generali** a supporto degli spoke del CN e potenzialmente di altri centri nazionali → **al Sud**
 - **5M€ previsti per hardware Cloud services ISO-certified** a supporto degli spoke del CN e potenzialmente di altri centri nazionali → **al Sud**
- Da TeRABIT (solo hardware):
 - **3.3M€ al sud (20%), 12.9M€ al nord (80%).**

INVESTIMENTI INFRASTRUTTURA		TOT	%SUD	TOT SUD	
	CINECA	Leonardo Upgrade	32.50	0%	0.00
	CINECA	Quantum computing	10.00	0%	0.00
	CINECA	Quantum accesso cloud	1.50	40%	0.60
	CINECA	Exascale	13.50	40%	5.40
	INFN	Tier 1 CNAF & Tier2 INFN	30.00	40%	12.00
	GARR	Garr T	25.00	94%	23.50
SPOKE	CMCC	Tier1 CMCC a Lecce	8.00	100%	8.00
	HUB	Tier1 Bicocca a Pavia (ENI)	1.00	0%	0.00
	HUB	Tier1 IIT a Genova	1.00	0%	0.00
	HUB	Tier1 UniPi a Pisa	1.00	0%	0.00
	CINECA	Tier 1 CNR/INAF al Tecnopolo	10.00	0%	0.00
	CINECA	Tier 1 Napoli CINECA/CNR	13.50	100%	13.50
	INFN	Tier1 UnivAq a L'Aquila	5.00	100%	5.00
	INFN	Tier 1 ESA a Frascati	5.00	0%	0.00
	INFN	storage/cpu ISO 27001 spoke MED/altri CCNN	5.00	100%	5.00
	INFN	storage/cpu altri spoke/altri CCNN	5.00	100%	5.00
	HUB	Tier1 SISSA	1.00	0%	0.00
	HUB	Tier1 Sapienza	1.00	0%	0.00
	HUB	Tier1 Enea	1.00	100%	1.00
SPOKE	UNIBO	Lab 1 Future HPC	2.00	0%	0.00
SPOKE	UNITO	Lab 2 Future HPC	1.50	0%	0.00
			173.50	46%	79.00

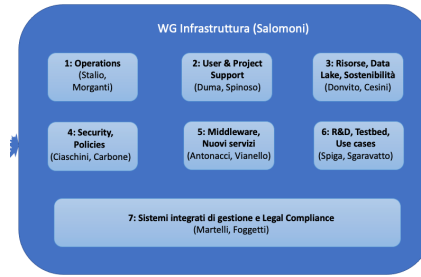
Vedi anche mail di gp del 30/3 su “indagine potenziamento centri di calcolo scientifico INFN”

→ dobbiamo rapidamente definire un capacity plan per storage, CPU, GPU, rete per fare partire il prima possibile programmazione e gare

Principi generali

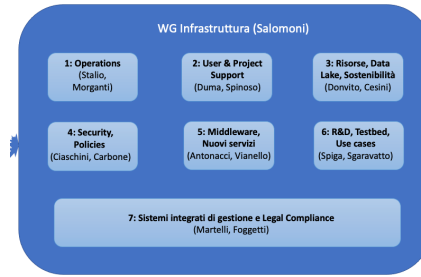
- Creiamo una **sigla nazionale per il WG infrastruttura**
- Il personale afferente al WG infrastruttura **rendiconta su quella sigla** con una percentuale P , **$50\% \leq P \leq 100\%$** , tenendo conto anche delle percentuali definite nei progetti PNRR.
- Consideriamo anche che **chi ad esempio già oggi lavora in un Tier-2** (magari con un cappello di esperimento) dovrebbe essere **comunque inserito nelle attività del WG**.

→ vediamo ora un po' di dettaglio per i vari WP proposti



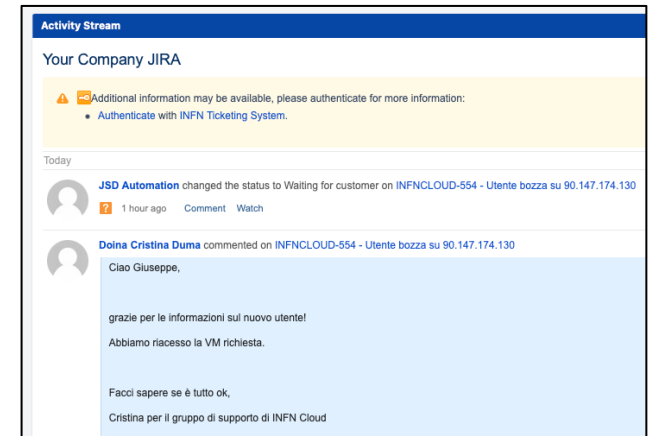
WP1: Operations

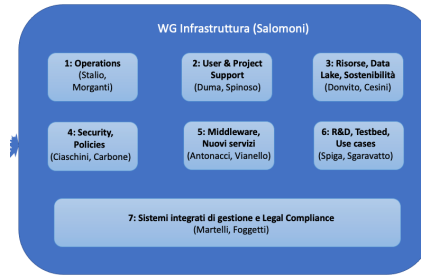
- **Monitoring** dello stato delle risorse, **accounting**, reportistica
- **Manutenzione e aggiornamenti** dell'infrastruttura
 - **Grid** (federazione di Tier-2) e **Cloud** (per backbone e in alcuni casi per le cloud federate)
- Gestione dell'**operatività e dei processi delle infrastrutture ISO** del datalake
- **Rapporti operativi con la rete GARR** (ad esempio per le soluzioni di virtual availability zones previste da TeRABIT)



WP2: User & project support

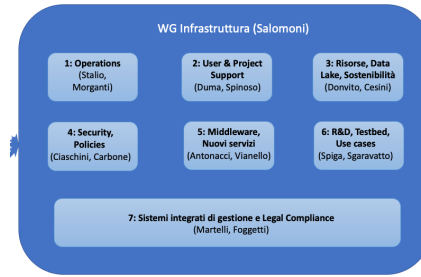
- **Supporto ai siti** (Tier-2, Cloud federate con INFN Cloud)
- **Supporto ai progetti** implementati sull'infrastruttura, inclusi quelli derivanti da proposte PNRR
- Supporto di **primo livello per tutte le risorse e servizi** (Grid e Cloud)
 - Il secondo livello è gestito a seconda del problema negli altri WP
- **Ticketing system, documentazione, knowledge base**





WP3: Risorse, data lake, sostenibilità

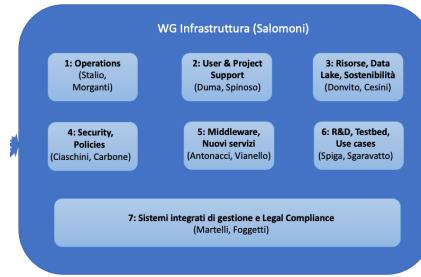
- Report a C3SN per validazione e referaggio gare o bandi legate ad acquisizione risorse. **Questo comprende sia hardware sia personale.**
- Definizione di **modelli di sostenibilità**
 - Relazione con le CSN, ad esempio per **utilizzo di risorse cloud vs. grid vs. local**
 - Relazione per quanto riguarda l'**uso delle risorse con i progetti nazionali ed europei, in collaborazione con il WG Progetti**
- **Data lake:**
 - Analisi dell'evoluzione dei Tier-x e del backbone di INFN Cloud
 - Evoluzione del modello architetturale di data lake
- Rapporto con **infrastrutture HPC** (in particolare CINECA), **in collaborazione con il WG HPC.**



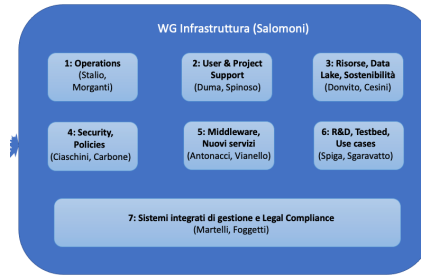
WP4: Security e policies

- Interfacciamento dell'infrastruttura di calcolo nazionale con **CSIRT nazionali e internazionali**
- Creazione e gestione di un «**SOC Infrastruttura**». Tra i suoi compiti: revisione e approvazione delle misure tecniche legate alla sicurezza informatica dell'infrastruttura
- Definizione delle **Rules of Participation** per l'accesso alle risorse
- Definizione delle **policies di sicurezza e di accesso alle risorse**, in collaborazione con il WP «Sistemi integrati di gestione e legal compliance»

WP5: Middleware, Sviluppo di nuovi servizi

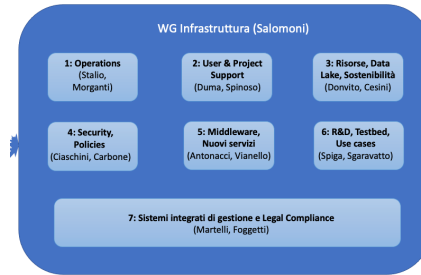


- **Supporto e sviluppo di:**
 - **Middleware:** IAM, PaaS, dashboard cloud, StoRM, VOMS, Argus
 - **Tool:** monitoring, accounting
- Sviluppo di **nuovi servizi a seconda dei casi d'uso** (attuale WP5 di INFN Cloud)
- Sviluppo di **soluzioni idonee alle infrastrutture ISO**, che spesso richiedono «bespoke solutions» o comunque particolari



WP6: R&D, testbed, use cases

- In collaborazione con il [WG su evoluzione tecnologica](#)
- Esempi di **testbed per il data lake**:
 - Test di implementazione tecnica ed evoluzioni. Ad esempio: replica automatica, utilizzo di Rucio, gerarchia di cache.
 - Integrazione di infrastrutture HPC
- Specificamente, **per infrastrutture ISO**:
 - Federazione di risorse e ingestion (anche in streaming) di dati distribuiti
- Quanto testato o sviluppato, **nel caso di passaggio in produzione**, passerà a WP1 Operations.



WP7: Sistemi integrati di gestione e Legal Compliance

- Pianificazione, organizzazione e conduzione del **sistema di gestione integrato** della Infrastruttura di calcolo INFN, che comprende l'insieme dei processi per garantire **Information security, Qualità dei servizi e Business continuity**
 - Nel campo della information security: gestione, in stretta collaborazione con il gruppo security, dei processi di definizione dei requisiti/obiettivi di sicurezza, risk assessment, audit interni ed esterni, incident management, misurazione della conformità e dell'andamento dei risultati raggiunti nel tempo.
- Definizione, in stretta collaborazione con il WP3 Security, delle **security policies**
- Supporto alla definizione di: **DMP, legal ed ethical requirement** per nuovi progetti e/o convenzioni, **in collaborazione con il WG Progetti**
- **Data protection**
- **Certificazione ISO** per i diversi siti e servizi INFN (a livello IaaS, PaaS o SaaS).
- **Da notare che questi temi non sono esclusivamente per i «siti ISO-certified».**
- Collaborazione con **Ufficio Trasferimento Tecnologico**

Prossimi passi

- **Formalizzazione del WG, del suo mandato, delle persone afferenti e delle sigle relative.**
- Ci sono **attività urgenti**, come il **capacity planning, l'impostazione delle gare in ottica PNRR, la definizione del sistema di recruitment.** Queste attività dovrebbero cominciare **adesso**.
- **All-hands meeting:** dopo l'estate? In collaborazione con la prima riunione del Forum del Calcolo?
 - (teoricamente si potrebbe pensare anche al Workshop sul Calcolo INFN di Paestum, ma penso che oltre a necessitare di una congrua preparazione, un AHM di questo WG richieda almeno una giornata)