Updates on Knowledge Transfer & Innovation Activities

CTSC Review, October 2020

Davide Salomoni, Cristina Vistoli

Knowledge Transfer strategy – Summary

- Knowledge Transfer continues to be key in order to:
  - Maintain and expand state-of-the-art know-how in technologies and processes with both academia and industry (KT is always both ways)
  - Strengthen connections with the surrounding environment (→ Tecnopolo)
  - Open new opportunities that are both strategic and technical and that normally go well beyond CNAF.
  - Provide sustainability regarding funding and manpower.
- We are greatly fostering the connection between the “PETT” and “SDDS” departments
  - See the next slides and the updated CTSC report.

KT: main lines of action

- Connection with Industry and Academia
  - E.g. Bi-Rex (→ Tecnopolo), SUPER (regional), Digital Innovation Hubs (regional, national, eventually European)
  - Recruitment
- Key Digital Technologies
  - E.g. sensitive data (with applications in multiple domains), IoT, Big Data Analysis
  - IEC/ISO 27001 certification – initially a CNAF-only initiative, now an INFN framework (with the vision of offering INFN Cloud – based certified services)
- E-health
  - Latest examples are COVID-related calls and opportunities, connections with Associazione Italiana di Fisica Medica
  - Alleanza Contro il Cancro, Harmony / Harmony Plus, Istituti Ortopedici Rizzoli, further submitted proposals such as ACTION, Italian Special Research Funding projects (FISR), other oncology-related projects, etc.

Innovation Strategy – Summary

- It originates from CNAF and has impact within CNAF, but it radiates toward all INFN (as per the official CNAF mandate)
- Continued leading role in INFN, national and international projects
  - Both as technical and strategic lead
- Expansion toward multi-disciplinary INFN-wide activities (INFN “Gruppo 5”)
  - QC, ML, COVID-19 & Environment, data libraries, data lakes, cultural heritage

Examples

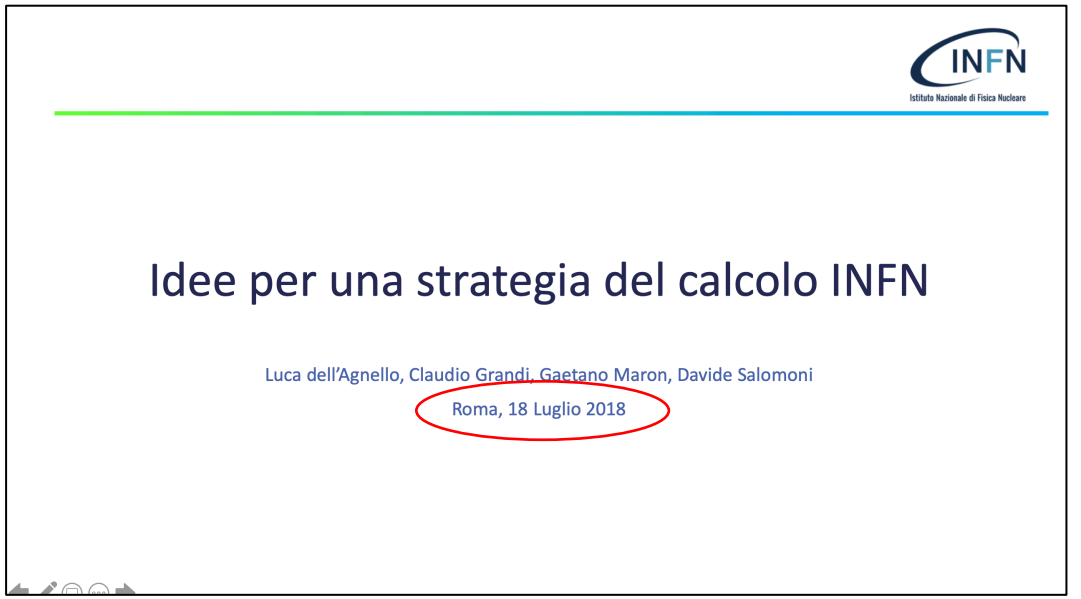
"The Committee believes that both the data centre operation unit and the user support unit are short of permanent human resources to cover the very vigorous program for future activities."

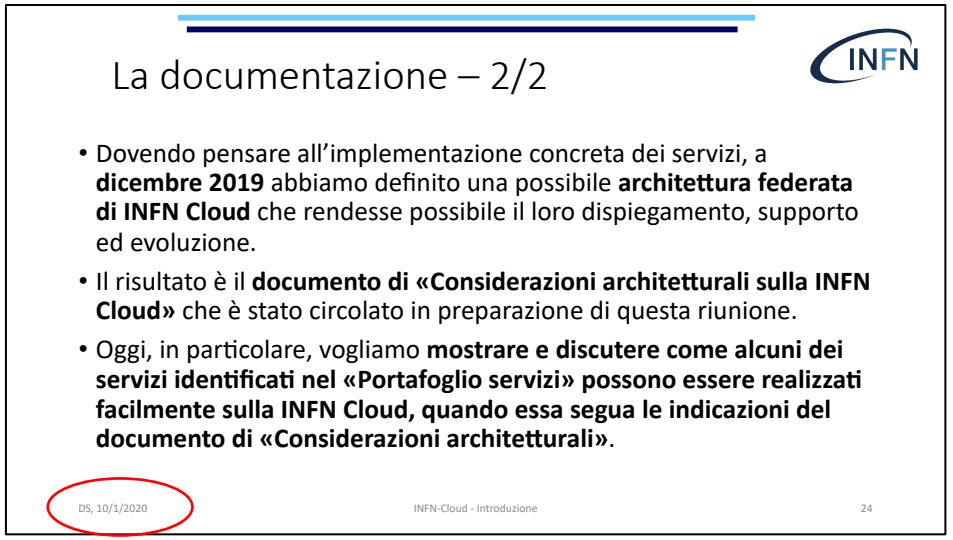
- XDC & DEEP are two EU projects (INDIGO-DataCloud follow-ups) that just ended.
  - **"Project has delivered exceptional results with significant immediate or potential impact"** (Final EC report for XDC, similar for DEEP).
- What did they bring us?
  - INDIGO-IAM is now the solution of choice for AAI in the entire WLCG (plus others, e.g. EOSC-Pillar, ESCAPE, STFC, etc.)
  - The INDIGO PaaS Orchestrator is the dynamic Cloud orchestration of choice in INFN Cloud, it is in the blueprint for the Italian Cloud Infrastructure, it is being integrated with Rucio for integrated data orchestration, etc.
  - The XDC Caching developments are already key in WLCG data lake activities
  - The INFN Cloud Dashboard inherits the experience of the DEEP dashboard
  - Regarding **manpower**: just XDC, DEEP & EOSC-hub have brought at CNAF 5 new people for 3 years (3 at the computing center, 2 at SDDS). 4 of them now have permanent positions, which would have not been possible without these projects.

INFN Cloud

"Details about INFN Cloud project. Referred several times in the answers to previous review recommendations but no much details in the 2019 report."

- It answers the **explicit request by the INFN President** to create and make rapidly available a Cloud strategy within INFN, exploitable also outside INFN.
- It pursues the **strategic objectives for INFN computing** that were identified already back in 2018 and since then confirmed by many national and international initiatives.
- **It continues, integrates and enhances experience, solutions, know-how and use cases** present and developed within INFN.



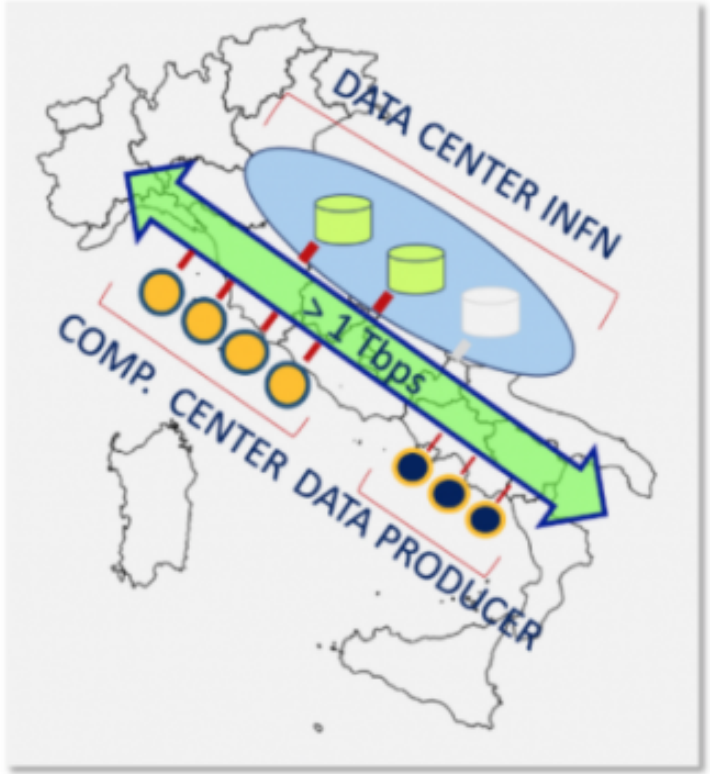


INFN Cloud architectural foundations

1. **Open source, vendor neutral architecture**, leveraging many years of INFN leadership, investments and know-how in e-infrastructures and distributed computing projects, implementing **extensible service composition**.
2. **Federation of existing Cloud infrastructures** for both compute and data.
3. **Consistent authentication and authorization technologies and policies** at all Cloud levels (IaaS, PaaS, SaaS) via OAuth and OpenID-Connect, supporting also legacy AAI solutions, via **INDIGO-IAM**.
4. **Dynamic orchestration of resources via the INDIGO PaaS Orchestrator** across all participating Cloud infrastructures, according to agreed SLAs and Rules of Participation.

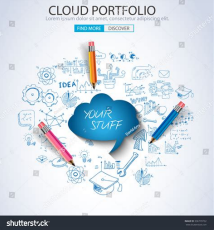
The INFN Cloud architecture

- **An INFN Cloud backbone** spanning the two main INFN computing sites (CNAF and Bari).
  - In each of these two sites there is an “INFN Cloud backbone infrastructure”, connected at high speed with each other.
  - The backbone is used to host the INFN Cloud core services, such as the PaaS core, the internal DNS, the logging and monitoring services, as well as user services that leverage backbone features, such as automated replication of object storage data across the two sites.
- **A set of distributed, federated cloud infrastructures** connecting to the backbone. Currently, the cloud infrastructures at CNAF and Bari (which are *not* the corresponding backbone infrastructures) are already connected to the INFN Cloud backbone, with several other INFN sites in the pipeline.



The INFN Cloud Organization

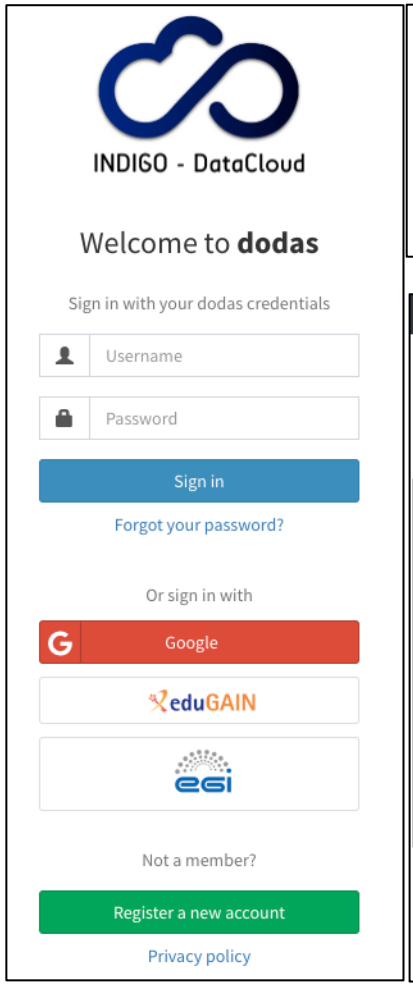
- INFN Cloud is internally organized into **5 Work Packages**, run by people belonging to several INFN sites in a fully distributed way:
  - WP1: Architecture, Operations and Service Portfolio
  - WP2: Documentation, User Support, Communication and Training
  - WP3: Monitoring and Accounting
  - WP4: Security, Policies and Rules of Participation
  - WP5: Service Evolution and New Developments
- The **INFN Cloud Management Board** is composed by the INFN Cloud Coordinator (Davide Salomoni, davide@infn.it) and by the WP Leaders. The reporting line is directly to the INFN Computing Coordination Committees and to the INFN President.
- The **current operational state** of INFN Cloud is pre-production but already serving several INFN experiments and collaborations, with full production state and general availability expected by early 2021, in sync with already planned resource expansion and connection of additional federated Clouds.



Which services does INFN Cloud offer?

- They **include**, for example:
  - Creation of VMs with different flavors and sizes.
  - Creation of containers (specify the container name) or of applications via docker-compose files.
  - Building blocks “as a service” for example for container orchestration (e.g. creation of a Mesos cluster or of a Kubernetes cluster as a service).
  - Pre-configured environments for example for data analytics (e.g. using ElasticSearch and Kibana or Spark).
  - Nonvolatile, object storage and Posix-compliant virtual file system solutions transparently connected to higher-layer services (e.g Jupyter notebooks as a service with permanent, replicated storage).
  - Dynamic clusters tailored to specific experiments (e.g. an automated full HTCondor installation realized on a k8s cluster, or a GPU-based Machine Learning-optimized environment).
  - Services leveraging transparent user-level encryption of disk volumes.
- **The service portfolio can be easily extended** with the simple addition / modification of TOSCA templates.

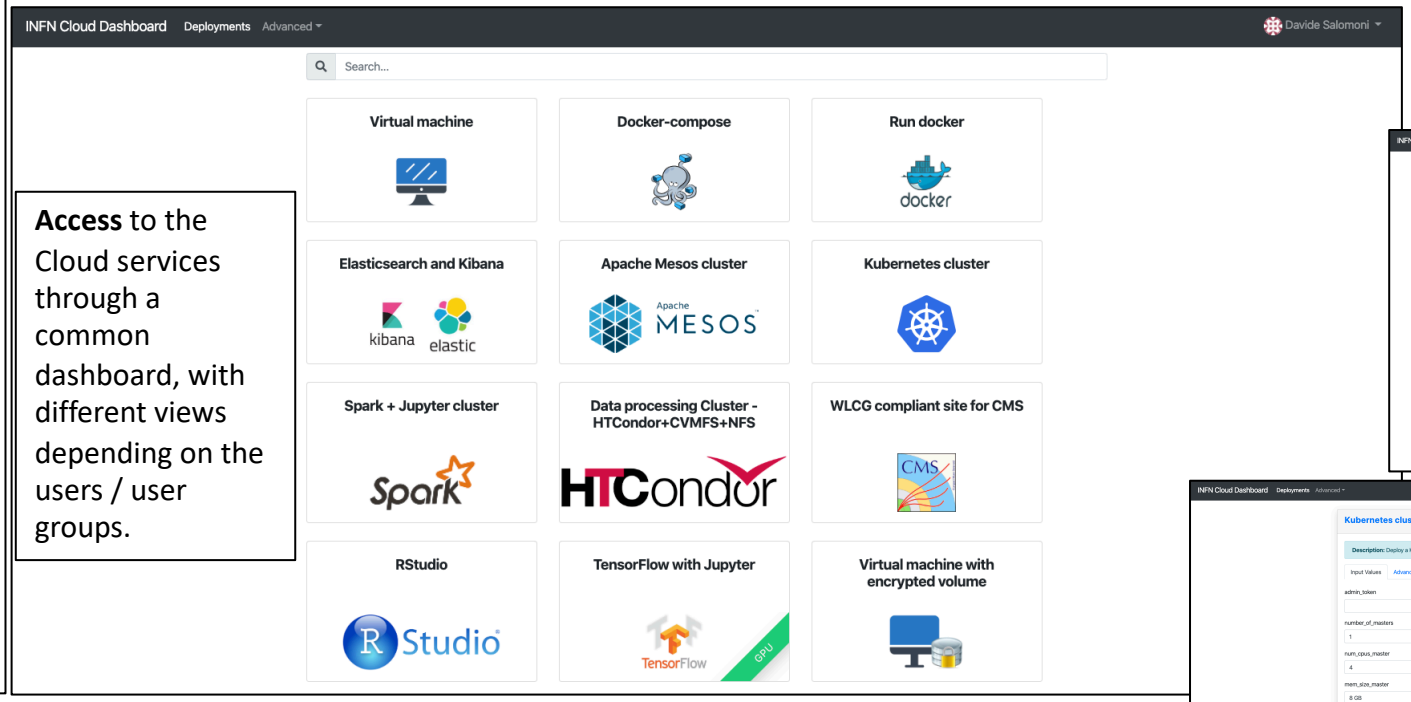
The INFN Cloud Dashboard



Authentication *can* be enabled for:

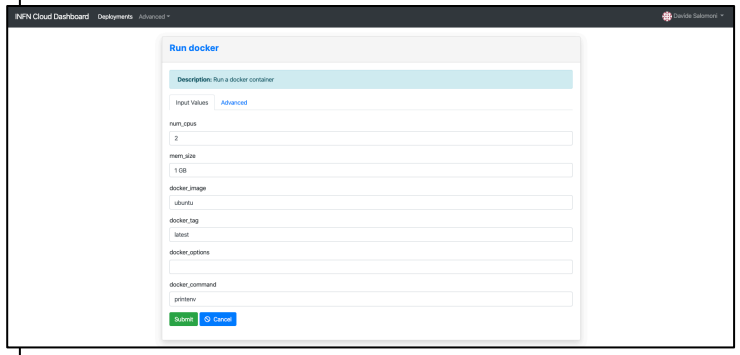
- Local username/password
- Google accounts
- EduGAIN (e.g. University, research centers, etc.)
- Other OIDC providers

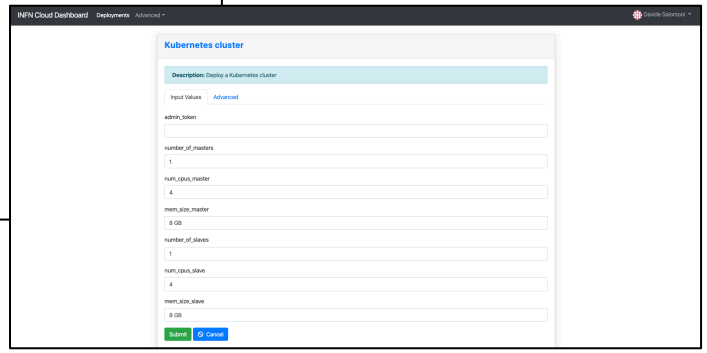
Transparent, multi-site **federation** for users of Cloud resources belonging to INFN and/or to other Cloud providers (private or public)

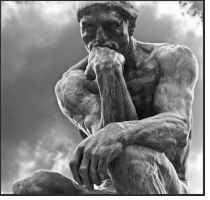


Access to the Cloud services through a common dashboard, with different views depending on the users / user groups.

Composed, high-level services easily customizable and configurable directly by users

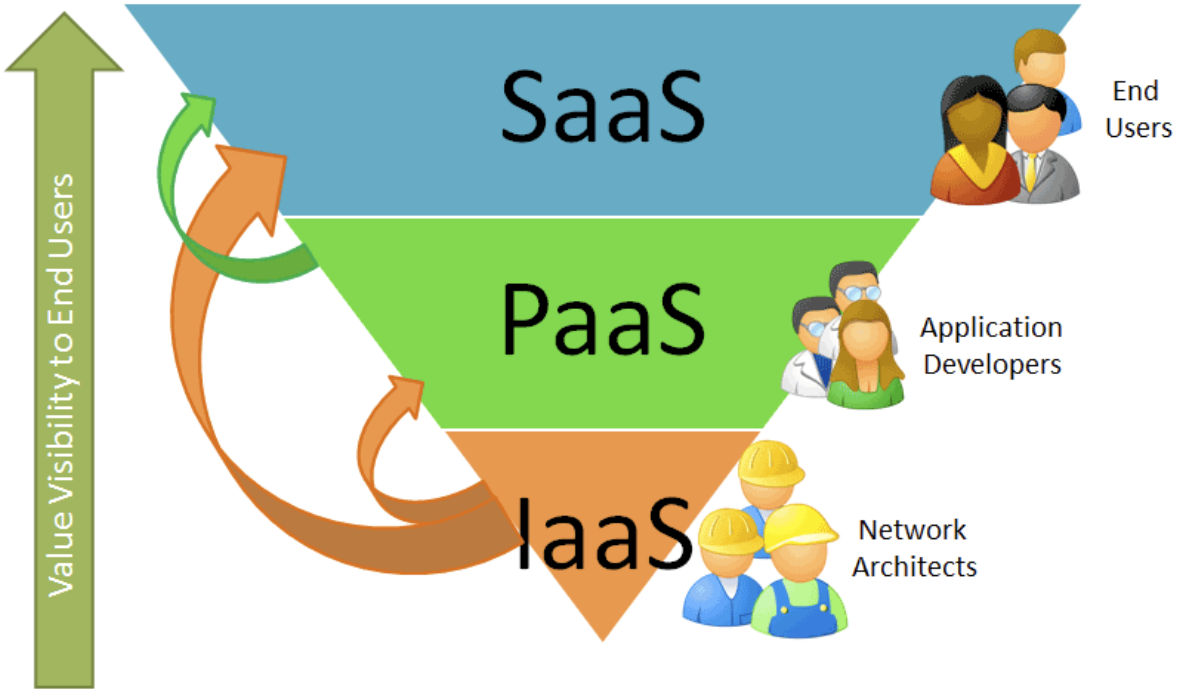


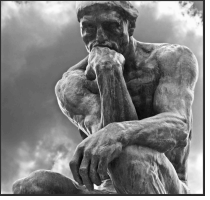




INFN Cloud Strategy (1)

- **What makes INFN Cloud special** is not IaaS. It is the ability to provide complex, composable, federated services. This is the consolidated trend in the Cloud world since several years, and it is what we have been working specifically at CNAF since INDIGO-DataCloud (2015).
- Strategically, therefore, INFN Cloud typically focuses on **Cloud levels > IaaS**.
- This of course does not mean that we ignore the IaaS, but it means that we want to focus on providing strong added value and support to INFN users.





INFN Cloud Strategy (2)

- **Inside**
  - Production data lake services beyond WLCG, reference for Cloud-based INFN services for the upcoming years
  - Testbed for porting apps to the Cloud, new computing models & services, training
- **Expansion toward sensitive data handling**
  - In e-health (see ISO/IEC 27001 compliance, a unique asset within INFN and in the academic & research Italian community), but also in IoT & Blockchain
- **Outside**
  - Integration with HPC resources (→ Tecnopolo, SUPER)
  - National & International exploitation
  - Connection with WLCG WG (e.g. DOMA, AuthZ) & CERN-CNAF.
  - Connection with public providers (e.g. AWS, Google)
  - Already now, **INFN Cloud technologies are in all proposals** recently submitted by INFN on distributed computing (including the big PON tenders in Southern Italy)

Strategy in Italy

- CNAF is a prominent part of the newly constituted INFN Group nominated by the INFN President to **propose changes to the INFN computing structure**.
- CNAF staff is part of the Italian group of experts for the definition of the Italian **National Research Program 2021-2027**.
- We promote **INFN Cloud as the blueprint for the future Italian National Cloud** at the ICDI (Italian Compute and Data Infrastructure) level.
  - We are also in the ICDI WG on Competence Centers and in those defining the ICDI Catalogue of Services.

Strategy in EU Projects

- **Success rate in the latest EINFRA calls: 100%**
  - INFN participated to 4 EINFRA calls; CNAF followed all of them in the preparation phase; all 4 were selected for funding. This is quite a unicum in the EU landscape.
  - For INFN: 661K€ (EGI-ACE, computing), 300K€ (DICE, data), 132K€ (C-SCALE, Copernicus), 312K€ (EOSC-Future, EOSC) → total 1.4M€ (for INFN; at CNAF: TBD)
- Key points:
  - Strategic link to the EOSC
  - Coordination within INFN
  - Careful selection of the Consortia
  - Technical exploitation of most of the INDIGO-derived outcomes
  - Due to a new EC tender strategy, funding is now often shifted toward service procurement (→ this includes funding of hardware resources)

Recruitment

"This unit appears to be short of human resources to cover the very vigorous program for future activities which are considered vital: the unit is experiencing difficulties for the recruitment of young professionals."

- Innovation activities are bringing us many Research Associates (A/R), who work in projects as well as in other areas (such as data center operations)
  - A/R that are currently recruited or in the process of being recruited: Harmony (x2), Smartchain (x2), We-Light, Fortress, Istituti Ortopedici Rizzoli, European projects (x4)
  - Strong interaction with User Support and its A/R team, with joint meetings and activities wherever possible.
  - Funding for other A/R is already available – we just wait to issue more grants to avoid having too many open positions.

Interactions with Academia, CS, etc.

"The Committee recommends to strengthen the ties with the Physics and the Computer Science departments of the local universities to attract students."
"CNAF does not seem to have much interactions with emerging data sciences."

- Besides the already mentioned collaborations in various projects and initiatives, we have been substantially focusing on giving courses at the University level. CNAF personnel currently teaches these courses:
  - Introduction to Big Data Processing, @ Bioinformatics + PhD course
  - Infrastructures for Big Data Processing, @ Bioinformatics + PhD course
  - Biomedical Data Bases, @ Bioinformatics
  - Programming, @ Physics
  - Python for Analytics, @ Statistical Sciences

Master's Degrees etc.

"The Committee encourages CNAF Management to study an alternative model for first line user support with a framework which includes projects largely manned by students."

- Besides having had 2 people who already graduated in 2020 with CNAF supervisors, we currently have 8 people doing their MD thesis with us. Topics:
  - Porting apps to the Cloud, IoT edge & mobile apps, data management, dynamic HTCondor, dynamic FaaS, heterogeneous computing software development, software code quality, prediction of the evolution of pandemic events.

Project Selection

"Currently there is no formal procedure to accept new projects and to evaluate the potential additional load on the structure."

- *"More details about the selection mechanism to accept new projects"*
  - This is a new procedure, that we have implemented starting this year. When the opportunity to participate to a proposal arises, this is first mentioned to CNAF Senior Management (the CNAF Director and the managers of the 3 CNAF Functional Units).
  - The proponent then writes a general description of the potential project. An internal referee is nominated, whose task is to fill in a "Preliminary Project Evaluation Template" (developed by us), which ranks the project for the 3 criteria of Excellence, Impact and Implementation, according to various pre-defined metrics.
  - These templates are then evaluated by CNAF Senior Management in order to provide guidance and feedback to the project proponent.
  - Note that this process is not always entirely applicable. While the first step (discussion with CNAF Senior Management) is mandatory, there are cases when timing or strategic opportunities may require a faster turnaround in the decision process.

Example of a Preliminary Project Evaluation

Criterio 1 - Eccellenza
Soglia minima: 3
Valutazione di questo criterio: 4,5

Vanno tenuti in considerazione i seguenti aspetti:

- Chiarezza e pertinenza degli obiettivi con le attività del CNAF.
- Solidità del concetto della proposta e credibilità della metodologia proposta.
- Quanto la proposta va oltre lo stato dell'arte e dimostra un potenziale di innovazione (per esempio: obiettivi all'avanguardia, nuovi approcci o concetti, nuovi prodotti o servizi).
- Quanto la proposta ha un approccio interdisciplinare che la possono rendere di uso potenzialmente applicabile a diversi domini scientifici.

La proposta risulta essere in linea con le attività di tipo progettuale e di sviluppo in seno al CNAF (IoT e progetti regionali, IoTwins, sembra voler riutilizzare gli outcom
Il CNAF si potrebbe anche avvalere
piattaforma sia per un uso interno a
o terze parti) delle risorse ICT.
La proposta risulta ben articolata

Criterio 2 - Impatto
Soglia minima: 3
Valutazione di questo criterio: 4

Vanno tenuti in considerazione i seguenti aspetti:

- Se e quanto l'impatto del progetto è menzionato nella proposta. L'impatto può essere visto come ampliamento della capacità di innovazione, creazione di nuove opportunità, miglioramento del posizionamento dell'INFN in generale e del CNAF in particolare, efficacia nei confronti della terza missione (impatto sulla società).
- Quanto i risultati del progetto possono contribuire concretamente ad avere l'impatto previsto nella proposta.
- Qualità delle misure proposte per sfruttare e comunicare i risultati del progetto a differenti tipi di pubblico.

La proposta copre diversi aspetti in termini di impatto:

- Enfasi delle soluzioni innovative per i partner tecnologici,
- Aspetti tecnici per i partner di progetto titolari degli use case,
- Impatto per la società, anche in termini finanziari, e riferiti agli use case.

La proposta sembra avere potenziali componenti innovative ma non risulta proprio chiaro come
ogetto. I KPI risultano coerenti con l'attività di progetto.
'impatto
nento agli o
a quella sta
ova la valo

Criterio 3 – Qualità ed efficienza dell'implementazione del progetto
Soglia minima: 3
Valutazione di questo criterio: 4

Vanno tenuti in considerazione i seguenti aspetti:

- Qualità ed efficienza del programma di lavoro, compresa una valutazione della quantità delle risorse CNAF previste (di personale e di materiale) e di quanto esse sono in linea con gli obiettivi del progetto.
- Appropriatezza della struttura di gestione del progetto, sia globale sia al CNAF, compresa la gestione dei rischi (che cosa succede se qualcosa va storto nell'implementazione) e dell'innovazione (cosa prevede di fare il progetto perché i suoi risultati siano realmente innovativi).
- Quanto il progetto coinvolge personale di più Unità Funzionali CNAF, personale di altre sedi INFN o personale di altri enti, al fine di avere tutte le competenze necessarie per la realizzazione degli obiettivi previsti.
- Appropriatezza nell'allocazione dei task all'interno del CNAF, considerando quanto tutti i partecipanti previsti al CNAF abbiano un ruolo adeguato all'interno del progetto e tenendo conto della reale possibilità di allocazione delle risorse interne.

In base al prospetto simulato (circa il 45-55 % del budget totale al CNAF) vengono previste per il CNAF 9 persone di cui:

- 7 staff
- 2 AR
- https://docs.google.com/spreadsheets/d/1DYJ2bYzvUz15aJKMfBJDKJq1iaXJBrYMugN6TvwG_jc/edit#gid=696967258

Il progetto deve necessariamente coinvolgere personale di diverse UF in seno al CNAF in quanto i task che vedono il possibile coinvolgimento del CNAF hanno diversa natura.
Competenze richieste (in giallo quelle di possibile interesse per il CNAF)

- WP1
  - Project management
- WP2
  - Privacy/security
- WP3 (INFN leader)
  - User requirements
  - Reference architecture design
  - Dissemination and standardization
- WP4 (INFN leader)
  - Paas orchestrator
- WP5
  - Software quality, release and maintenance
  - Pilot

SDDS

It is not clear how the outcomes and deliverables of R&D activities are evaluated for CNAF.

- *"SDDS continues to be very successful*
  - *What is the long-term vision?*
  - *What are the synergies with other activities at CNAF?*
  - *Is there a 'project office' to follow up calls and reports and insure cohesion?"*

Vision

- SDDS is increasingly connected locally to PETT and to the Data Center for the exploitation of project outcomes, to continue creating strategic innovation and opportunities and to support and develop key software components. But it is also recognized INFN-wide as the place where most innovation initiatives related to INFN Computing take place.

Synergy

- Since several months, we now have a weekly "Projects Meeting" to which all CNAF departments participate, to discuss progress and blocks for all the CNAF-related projects.
- We also have periodic CNAF-wide meetings to present new approved projects and to report on concluded projects (example: https://agenda.infn.it/event/21246/).

Projects

- Strategically, projects are discussed with CNAF senior management (see the answer on project selection) as well as with the INFN Computing Committees, since projects follow up is often an INFN-wide decision, with global impact.
  - Often, strategic leadership is kept at CNAF, with involvement of several INFN sites in both project preparation and operational duties (latest example: the 4 EINFRA proposals recently approved).

CNAF and the EOSC

- *"Can I understand CNAF's position as a technology driver about the participation and future role in EOSC, is CNAF mandated for Italy in the Association?"*
  - CNAF is part of INFN. INFN participates as founding member to ICDI, which is part of the EOSC AISBL. We at CNAF participate in most of the technical and strategical initiatives related to the EOSC, either directly or indirectly.
  - See above for the role of CNAF in the EOSC-related calls. As another example, INFN (not CNAF) is the member participating to EGI – however, we at CNAF do participate to the EGI activities (e.g. in EGI Executive Board) and thus try to steer – in agreement with the INFN President and the INFN Computing Committees – the EOSC strategy.