



Riassunto sulle call EU per attività IT/ReCaS

Giacinto DONVITO
INFN-Bari

Agenda



- Progetti H2020:
 - INFRAEOSC-07-A1 (Servizi di Compute)
 - INFRAEOSC-07-A2 (Servizi di Data Management)
 - INFRAEOSC-07-A6 (Servizi di supporto alla Comunità Earth Observation)
 - INFRAEOSC-03 (Implementazione della EOSC)
 - DT-TDS-04-2020 (AI for Genomics and Personalised Medicine)
- Considerazioni sulle attività di R&D
- Considerazioni economiche

H2020: INFRAEOSC-07-A1 (Servizi di Compute)

- Deadline: 18 June 2020 (**8M€**)
 - (a1) Distributed and cloud computing resources enabling researchers and other users to process and analyse data in a distributed computing environment. The services should include, but not be limited to, **running virtual machines** on demand with complete control over computing services, **executing compute and data-intensive workloads**, analysing large datasets and executing parallel computing tasks, utilizing large amount of processing capacity over long periods of time, sharing resources and enabling collaborative research.
- Storia travagliata, ma alla fine d'accordo con il management dell'ente, si è deciso di andare con la cordata di EGL.
 - Ci sarà una cordata parallela e concorrente a cui partecipa il CINECA
- **Sedi coinvolte: Bari, CNAF, Perugia, Torino**
- Noi (Bari) abbiamo chiesto di partecipare per:
 - il supporto alla INDIGO PaaS (servizio che è diventato core sia per ReCaS che per INFN-Cloud)
 - Il supporto alle comunità di utenti fornendo risorse hw, che vengono ripagate con mesi uomo (staff) ma anche con Unit Cost, che possiamo rendicontare e incamerare per le nostre esigenze
- Per **Bari** si potrebbe arrivare forse a **~200k€ di finanziamento in 30 mesi**
 - **Partenza prevista circa Gennaio 2021**
- Da rendicontare praticamente tutto con **personale staff**

H2020: INFRAEOSC-07-A2 (Servizi di Data Management)



- Deadline: 18 June 2020 (7M€)
 - (a2) Data services providing cost-effective and interoperable solutions for data management and long-term curation and preservation. The needs for discoverability, accessibility, interoperability, text and data mining, transfer of data between data resources and computational facilities, storing, managing and accessing persistent identifiers (PIDs) and essential metadata (PID records), reusability, annotation and integrity across disciplines should be accommodated in multi-disciplinary and multi-sectoral contexts. All these services should be aligned with the EOSC rules, considering also ethical and regulatory requirements for sensitive data.
- Per espressa richiesta del Management stiamo partecipando al proposal con il CINECA e la compagine di EUDAT/PRACE.
- **Sedi coinvolte: Bari, CNAF, ?**
- Su questa call, ci aspettiamo che il ruolo di «attore principale» lo vorrà fare il CNAF
- Per **Bari** si potrebbe arrivare a **<100k€ di finanziamento in 3 anni**
 - Per implementare un po' di tecnologia che padroneggiamo noi. Non ci aspettiamo che il CNAF ci lasci rendicontare risorse HW, ma vedremo...
- Da rendicontare praticamente tutto con **personale staff**
- **Partenza prevista circa Gennaio 2021**

H2020: INFRAEOSC-07-A6 (Servizi di supporto alla Comunità Earth Observation)



- Deadline: 18 June 2020 (**2M€**)
 - (a6) Additional research enabling services that are not covered under any of the areas described in areas a1 to a5 and that do not overlap with existing services accessible through the EOSC Portal. Services should be of a generic value to the whole research community, such as text and data mining or Copernicus services.
- Vista la forte presenza di una comunità di EO a Bari abbiamo risposto positivamente alla richiesta di partecipare a questa call, promossa da EGI
- **Sedi coinvolte: Bari, CNAF?**
- Per **Bari** si potrebbe arrivare a **~150k€ di finanziamento in 3 anni**
 - «Vendere» la tecnologia che usiamo noi anche nella comunità di EO. Supportare la comunità con risorse HW. Cercare di portarci a Bari un po' dei dati di EO
- Da rendicontare praticamente tutto con **personale staff**
- **Partenza prevista circa Gennaio 2021**

*few chances
of success*

H2020: INFRAEOSC-03 (Implementazione della EOSC)



- **Deadline: 18 June 2020 (40M€)**
 - Continuare l'implementazione della EOSC. Sarebbe il successore di EOSC-Hub
 - È verosimile che questa volta l'esposizione di INFN e di INFN-Bari sia minore che nel progetto precedente
 - Abbiamo deciso di non esporci troppo su ruoli di coordinamento (viste le altre attività in corso)
- In questa call ci sono tutte le infrastrutture (EGI, EUDAT, OpenAIRE, etc) e tutte le Comunità ESFRI (Elixir, Lifewatch, Physics, etc)
 - Quindi anche se meno esposti che in passato, dobbiamo esserci per forza.
- **Sedi coinvolte: Bari, CNAF, Torino, Perugia, Padova, ?**
- Per **Bari** si potrebbe arrivare a **~150/200k€ di finanziamento in 36 mesi**
- Da rendicontare praticamente tutto con **personale staff**

H2020: DT-TDS-04-2020 (AI for Genomics and Personalised Medicine)



- Deadline: 18 June 2020 (**10M€**)
 - Proposals should develop and test AI solutions for linking genomics repositories across the EU, including banks of "-omics" and health related data, biobanks and other registries (including e.g. rare disease registries), with the view of supporting clinical research and decision making. By combining sequenced genomic data and other medical data, physicians and researchers can understand better diseases at a personal level and can determine the most appropriate treatment for a particular person. The focus should be to reduce the burden of diseases for which a treatment exists and to apply such treatments in a more targeted way, to identify new evidences on the predictive value of the AI solutions and to enhance the diagnostic capacity e.g. for rare or low prevalence and complex diseases.
- Ci è stata chiesta dalla JRU Elixir-IT la partecipazione a questa call con due proposal
- **Sedi coinvolte: Bari**
- «Vendere» la tecnologia che usiamo noi, nella comunità della «Personalised Medicine»
- Prepararci a supportare questi use case per le comunità locali (gestione di dati sensibili, e supporto agli algoritmi di AI)
- Coinvolgimento del GR 5 per la parte algoritmica di AI
- Da rendicontare sostanzialmente con **personale staff, più eventualmente 1 assegno dedicato ad attività specifiche**

little chances
of success

Considerazioni sulle attività di ricerca

- Dalle attività dei progetti INDIGO-DataCloud e EOSC-Hub sono stati resi disponibili:
 - Una soluzione per l'uso di risorse opportunistiche (DODAS) già usata per CMS, e vari esperimenti di GR2
 - in vari casi: CINECA, HelixNebulaScienceCloud, Microsoft Azure, etc
- Dai progetti INDIGO-DataCloud e XDC:
 - È stata implementata la soluzione che è alla base di altri progetti EU che sostengono WLCG e nazionali IDDLS che è il progetto che ha l'obiettivo di implementare il data lake INFN.
- Dal progetto DEEP-HybridDataCloud:
 - Sono state realizzate le tecnologie che nella prossima versione del data center metterà a disposizione le risorse di GPU e framework per l'esecuzione di algoritmi di IA anche per i gruppi baresi.

Considerazioni sulle attività di ricerca

- Le attività cominciate nei progetti H2020 nel 2015 adesso stanno permettendo di costruire la «INFN Cloud» in cui Bari si è guadagnato un ruolo di primissimo piano
- Ben 3 persone con ruolo di coordinamento su 5 WPs disponibili
- L'unico altro sito oltre al CNAF: non è mai stato in dubbio
- Su questo lavoro si baserà l'implementazione della prossima infrastruttura di calcolo dell'Ente (modello data lake).
 - È importante mantenere un ruolo di primo piano se vogliamo che Bari sia ancora fra i data center importanti e mantenuti anche nel futuro
 - Bari potrebbe essere l'unico sito INFN diverso dal CNAF ad ospitare dati (long-term) e non solo CPU.

Considerazioni sulle attività di ricerca



- Dobbiamo farci trovare pronti sia tecnologicamente che come risorse ai nuovi sviluppi di WLCG e dei nuovi Computing Model degli esperimenti in futuro

Considerazioni di carattere generale



- Il gruppo IT non ha finanziamenti strutturali come sulle commissioni scientifiche
- È necessario per assicurarsi le attività di R&D e il sostentamento dell'infrastruttura di produzione (che comprende anche i Tier2) tramite attività con finanziamento esterno:
 - Call EU, Nazionali, Regionali, Convenzioni e commesse, etc
- La ricaduta di questi R&D è sempre stata riportata in produzione sia per gli esperimenti INFN che localmente per il mantenimento del data center.

Considerazioni di carattere generale



- Nell'ottica di dover supportare l'infrastruttura di ReCaS:
 - Manutenzioni (come avrete notato viene chiesto sempre meno supporto ai gruppi su questo capitolo)
 - Upgrade dell'HW
 - Supporto per l'implementazione dei PON
 - Spese non coperte dai PON
 - Piccole aggiunte (Rack, etc)
 - Supporto alle comunità di utenti (locali, e INFN in generale) con possibilità di fare AR dedicati
- È importante restare nel solco dei progetti finanziati non solo per mantenere il know-how allo stato dell'arte, ma anche per ragioni economiche
- Contando solo sul personale staff, possiamo «rendicontare» fino a circa 200k€/anno

Considerazioni di carattere generale



- Già ora il gruppo ReCaS sta contribuendo in vario modo alla Sezione e ai gruppi di ricerca di INFN-Bari e del DIF:
 - Sta per essere bandita una posizione di Coll. Amm. TD per dare una mano alla gestione dei progetti (per tutti) e delle Gare
 - Fornitura di servizi, supporto utenti e risorse:
 - Owncloud per rendicontazioni progetti, condivisione file, backup, etc,
 - Supporto al calcolo batch, anche parallelo e con uso di GPU,
 - Servizi Virtuali per: siti web, masterclass, Jupyter Notebook, Gitlab, CI, etc
 - Sviluppo di portali e interfacce grafiche per semplificare l'interazione con le risorse
 - Supporto alla sottomissione e esecuzione di progetti finanziati in cui l'infrastruttura di ReCaS può essere un importante valore aggiunto
 - Possibilità di accendere AR per persone di esperimento che vogliono cimentarsi attività più IT (dal supporto utenti alle attività nei framework di analisi moderni). Utile per acquisire conoscenze IT e arricchire il proprio CV.