

# Position paper INAF sullo sviluppo e sul supporto alle attività cloud

*Fabio Pasian – INAF – 30.10.2012*

## La comunità astronomica italiana e l'Istituto Nazionale di Astrofisica

La comunità astronomica italiana è attiva all'interno del sistema Universitario nazionale e, soprattutto, dell'Istituto Nazionale di Astrofisica (INAF), distribuito su 16 Strutture di ricerca (Osservatori Astronomici e Istituti ex-CNR) con una ventina di sedi; l'Ente supporta direttamente anche due sedi osservative al di fuori del territorio italiano, a La Palma (isole Canarie) e su Mount Graham (Arizona, USA). L'INAF ha circa 1100 dipendenti: di questi circa 700 sono ricercatori e tecnologi; oltre al personale di ruolo, circa 500 ricercatori sono associati a INAF e operano nelle Strutture dell'Ente con forme diverse di collaborazione: di questi, oltre 300 sono giovani con contratti a tempo determinato (TD, assegni e borse di ricerca).

Questa comunità partecipa attivamente e spesso occupa ruoli di coordinamento all'interno di grandi progetti strumentali, tra cui quelli identificati da ESFRI (CTA, E-ELT, SKA) e le missioni spaziali del programma "Cosmic Vision 2025" dell'Agenzia Spaziale Europea (Euclid, Solar Orbiter, JUICE), oltre a molti altri progetti non meno impegnativi scientificamente e per risorse impiegate. Per la realizzazione dei loro obiettivi scientifici, tali progetti affrontano notevoli sfide tecnologiche sia nel campo della costruzione strumentale, che in quello delle infrastrutture digitali. Infatti, tali progetti sono intrinsecamente data-intensive e pongono sfide computazionali di difficile soluzione.

## Infrastrutture informatiche INAF

In alcune delle sue Strutture, l'INAF dispone di infrastrutture di storage e di calcolo ad alte prestazioni, di dimensioni piccole o medie. In particolare, questi sistemi di elaborazione sono cluster paralleli, più alcuni sistemi basati su GPU e ibridi CPU-GPU. Alcuni di questi, per lo più finanziati dal MIUR tramite PON, sono interconnessi alle infrastrutture Grid italiana ed europea (EGI). Le risorse di calcolo sono utilizzate per scopi di ricerca interni all'INAF, per offrire servizi nell'ambito delle attività operative (missioni spaziali, telescopi a terra) e sono in parte usate dai dipartimenti/sezioni di astronomia delle università ospitati presso le Strutture dell'Ente, e anche come cofinanziamento in progetti europei e nazionali compartecipati dall'INAF. L'INAF non eroga servizi di calcolo ad alte prestazioni verso l'esterno se non nell'ambito di progetti di ricerca con partner esterni, ma usa le infrastrutture per i suoi scopi istituzionali e di ricerca.

I settori di attività nel campo dell'informatica applicata sono molto ampi e riguardano lo studio dei sistemi HPC e Grid/Cloud finalizzato alle tematiche di ricerca in astronomia e astrofisica, quali l'elaborazione di grandi quantità di dati osservativi, la produzione di modelli, la fluidodinamica, la gestione e l'archiviazione dei dati, il data mining e l'information retrieval. Queste attività hanno portato anche alla progettazione e realizzazione di sistemi software per il calcolo, librerie e ambienti di sviluppo ed esecuzione di applicazioni ad elevato parallelismo, da realizzarsi su sistemi ad alte prestazioni e/o tramite calcolo distribuito.

L'INAF ha inoltre un'attività di calcolo legata agli aspetti amministrativi dell'Ente, espletati presso la Sede Centrale e il CED Amministrativo (presso IRA-Bologna). Il coordinamento delle attività informatiche dell'Ente è stato svolto dal 2004 al 2011 dal Servizio Sistemi Informativi (INAF-SI) con sede presso OA-Trieste: INAF-SI ha supportato l'utenza (scientifica e non) nell'utilizzo delle infrastrutture informatiche, e ha coordinato le attività di ricerca e sviluppo in campo ICT se finalizzate a un futuro utilizzo come servizio all'utenza.

### **Attività INAF nel campo delle infrastrutture digitali multi-disciplinari**

INAF è attiva nel campo delle infrastrutture distribuite di calcolo e di gestione dati da oltre dieci anni, tramite la partecipazione a progetti finanziati dai Programmi Quadro dell'Unione Europea, sia tematici (in particolare dedicati alle osservazioni remote e alla presenza virtuale, e all'interoperabilità tra gli archivi di dati astronomici distribuiti a livello mondiale – il cosiddetto “Osservatorio Virtuale”) che multidisciplinari (relativi alla realizzazione di infrastrutture di calcolo distribuito e alle relative applicazioni all'astrofisica). Di particolare rilievo il ruolo ricoperto da INAF (chairman 2007-10) all'interno dell'International Virtual Observatory Alliance (IVOA) l'organizzazione mondiale per l'Osservatorio Virtuale, e delle iniziative internazionali per il calcolo distribuito (coordinamento della comunità Astronomia e Astrofisica in EGEE-III ed EGI).

Sia nel campo della gestione dei dati distribuiti geograficamente, sia in quello del calcolo distribuito, che nel crescente interesse presso la comunità scientifica per le tematiche relative al data preservation, INAF ha sempre dimostrato uno spiccato interesse a collaborare con le altre discipline scientifiche per la realizzazione di infrastrutture interoperabili integrate e l'implementazione e messa in produzione di servizi utilizzabili dalla comunità scientifica intesa nel senso più ampio.

### **Il ruolo di INAF all'interno di una infrastruttura digitale nazionale per Università e Ricerca**

La partecipazione INAF, e in genere dell'intera comunità astronomica, ai grandi progetti sopra menzionati (con particolare riferimento a quelli selezionati da ESFRI e da ESA Cosmic Vision) rende urgente la realizzazione e messa in opera di servizi multidisciplinari e geograficamente distribuiti, per poter rispondere nel modo più efficace alle sfide scientifiche e tecnologiche che questi progetti pongono. Oltre all'ovvia necessità di condivisione di risorse che portino a un'ottimizzazione delle capacità di calcolo e di gestione distribuita dei dati, ci sono svariati aspetti tecnologici che solo un coordinamento inter-Ente può realizzare in modo efficace; come semplice esempio si consideri la necessità di controllo sull'accesso alle risorse tramite autenticazioni federate SSO (Single-Sign-ON).

Le applicazioni di calcolo ed elaborazione dati della comunità astronomica sono ampiamente diversificate dal punto di vista computazionale e di data flow. Le necessità computazionali spaziano dall'uso di semplici cluster locali di fascia bassa a grosse infrastrutture costituite da risorse di calcolo eterogenee (grid, cloud, HPC). La soluzione tecnologica che è meglio rispondente a queste necessità è un ambiente integrato basato su una rete digitale ad alta efficienza che permetta all'utente di accedere in modo trasparente a risorse federate ed eterogenee (HPC-HTC-dati) tramite una infrastruttura digitale all'uopo costruita e mantenuta a livello operativo (“Cloud per la Ricerca”).

Dal momento che le necessità scientifiche e tecnologiche della comunità hanno evidenziato un elevato grado di dinamicità e di evoluzione, un requisito fondamentale che si richiede alla nascente infrastruttura digitale è una capacità di adattarsi facilmente alle nuove esigenze dell'utenza; essa deve essere quindi pensata in termini modulari per facilitarne la modificabilità ed espansione nel tempo.

Si ritiene pertanto indispensabile procedere a un coordinamento tra i Consorzi di supercalcolo, il Consortium GARR e gli Enti di Ricerca (almeno INFN, ENEA, INAF, CNR, INGV, OGS) allo scopo di realizzare la necessaria sinergia che permetta all'infrastruttura digitale multidisciplinare e geograficamente distribuita di essere realizzata e messa in opera tempestivamente.

A tal fine INAF ribadisce:

- la volontà di utilizzare per i propri progetti l'infrastruttura digitale esistente, propria e di altri Istituti, in modo coordinato allo scopo di ottimizzare l'utilizzo delle risorse;
- la volontà di potenziare centri anche non-INAF che garantiscano un'operatività continua, con risorse da condividere all'interno della futura infrastruttura digitale;
- la volontà di condividere risorse INAF esistenti, e da acquisire in futuro, all'interno di un sistema coordinato di Cloud per la Ricerca;
- la disponibilità a mantenere operative le risorse contribute all'infrastruttura digitale condivisa;
- la disponibilità a supportare con proprie risorse umane
  - lo sviluppo tecnologico della Cloud per la Ricerca secondo le necessità stabilite con gli altri partner;
  - il training del proprio personale di ricerca e tecnico finalizzato all'utilizzo e alla gestione, rispettivamente, dell'infrastruttura digitale condivisa;
  - il mantenimento e lo sviluppo di competenze specialistiche disponibili nei propri centri e la pianificazione concertata, in base alle esigenze nazionali, degli investimenti di potenziamento formativo (borse di dottorato, borse di studio, ...) per la crescita delle stesse.
- la volontà di pianificare in modo coordinato, in base alle esigenze comuni a tutti i partner, degli investimenti di potenziamento infrastrutturale e formativo;
- la volontà di partecipare, congiuntamente agli altri partner, alle iniziative nazionali ed europee nel campo, incluse le domande di finanziamento sui canali opportuni.