

# ROADMAP PER LA PARTECIPAZIONE DI INAF AL LOW FREQUENCY ARRAY (LOFAR)

G. Brunetti, F. Govoni

con il contributo del Working Group WG-F03-01

U.Becciani (INAF Osservatorio Astrofisico Catania)

P.Bolli (INAF Osservatorio Astrofisico Arcetri)

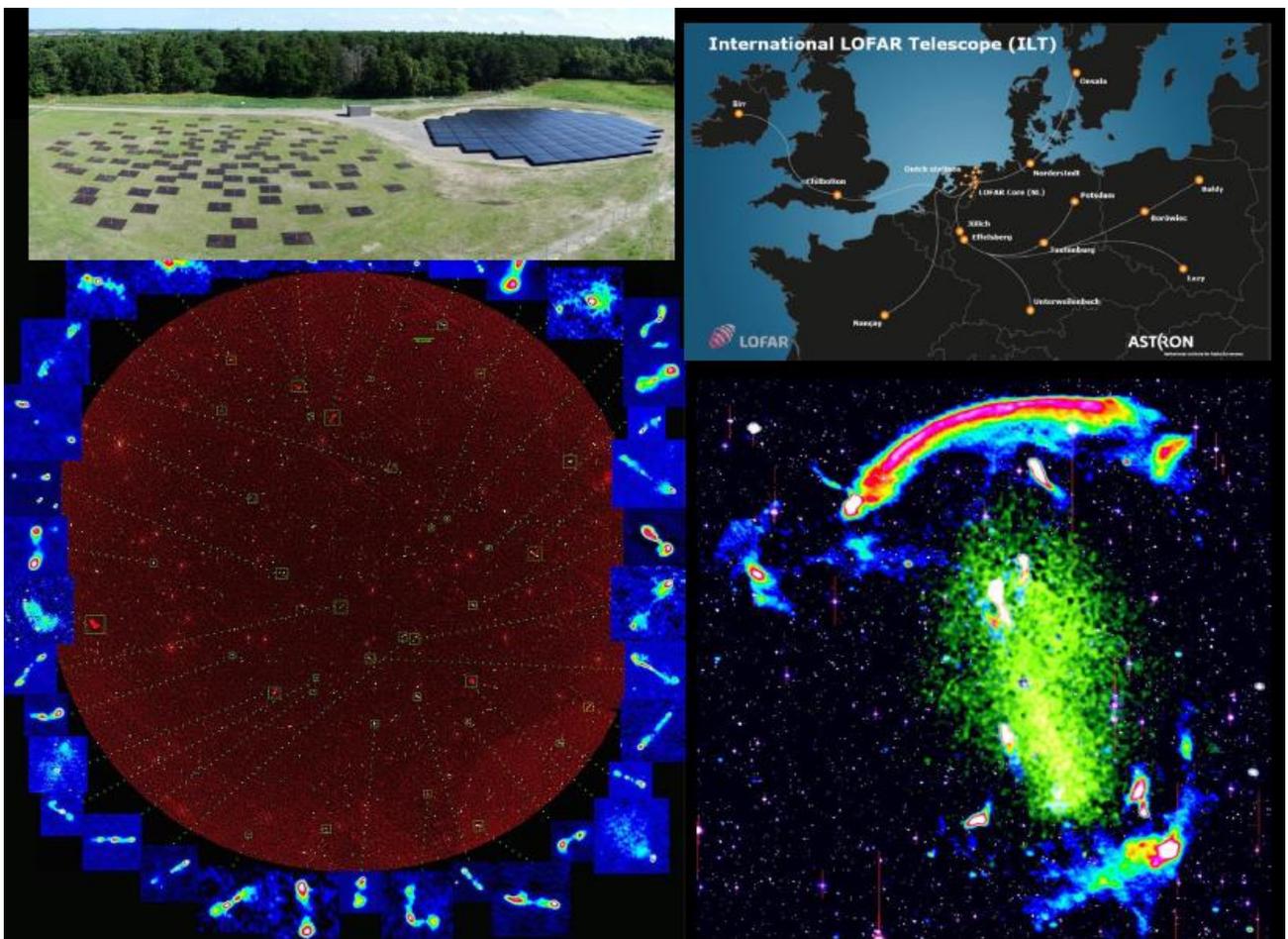
A.Bonafede (INAF IRA Bologna)

J.Monari (INAF IRA Medicina)

M.Nanni (INAF IRA Bologna)

F.Perini (INAF IRA Medicina)

G.Taffoni (INAF Osservatorio Astronomico Trieste)



## EXECUTIVE SUMMARY

La corsa verso SKA ha generato un insieme di precursori/pathfinders che oggi e per una decade guideranno la ricerca di frontiera in ambito radioastronomico. Non prendendo parte a nessuno di questi precursori/pathfinders l'Italia si precluderebbe una partecipazione con la dovuta massa critica a quella moltitudine di sviluppi scientifici e innovazioni tecniche che sono in continua evoluzione e che stanno preparando la strada all'esplorazione scientifica con SKA.

**LOFAR è il più grande precursore di SKA in termini di area efficace e di data-rate generato.** E' un aperture array rivoluzionario che si estende su scala Europea e che sta iniziando ad aprire una nuova finestra osservativa alle basse frequenze radio promettendo fondamentali passi in avanti in una moltitudine di ambiti dell'astrofisica e cosmologia.

Al momento i gruppi di ricerca afferenti a LOFAR sono strutturati in 6 Key Projects (KP) all'interno dei quali viene sviluppato gran parte del know-how tecnico e scientifico e che coprono numerose aree di ricerca alle quali la nostra comunità è fortemente interessata. La partecipazione italiana a LOFAR è stato un obiettivo inseguito da diversi anni dalla nostra comunità ma che ha trovato solo recentemente la possibilità di concretizzarsi grazie ad una iniziativa dei vertici dell'INAF per assicurare la partecipazione ad un precursore SKA science ready finanziata su fondi SKA Industria.

Nel Giugno 2017 INAF ha iniziato una negoziazione con l'International LOFAR Telescope (ILT) e ASTRON allo scopo di aderire al progetto LOFAR **prevedendo l'ingresso immediato nei KP del personale di ricerca.**

**Il negoziato si è concluso con una proposta di roadmap che garantirebbe ad INAF il ruolo di Full Member di ILT sin DA SUBITO, compreso l'accesso ai KP e una fetta di tempo di osservazione riservato all'anno.** Questo ruolo sarebbe garantito in una prima fase dal coinvolgimento tecnologico di INAF nell'upgrade previsto per LOFAR (LOFAR 2.0), e in una fase successiva (presumibilmente dal 2021-22) dall'acquisizione di una Stazione LOFAR 2.0.

In particolare lo schema previsto dalla roadmap è il seguente :

1. INAF guiderà un consorzio Italiano per LOFAR (LOFAR IT) di cui sarà rappresentante legale nell'ambito delle relazioni con ILT. Il primo passo prevede un consorzio formato da INAF e Dip.di Fisica dell'Università di Torino (UniTo);
2. INAF siglerà un accordo con AstroTec con il quale si impegna ad acquisire una Stazione LOFAR 2.0 da installare a Medicina (Bo) entro il 2021-22;
3. INAF aderirà ad un accordo di collaborazione tecnologica 2018-2021 con ASTRON per lo sviluppo di LOFAR 2.0;
4. LOFAR IT (INAF) si impegnerà a versare la quota annuale prevista per il consorzio ILT.

In base alle condizioni concordate con i partner (ILT, ASTRON, AstroTec), il costo di questa operazione per INAF è **stimabile in un massimo di 2.4 MEuro in 5 anni** indipendentemente dai dettagli del consorzio LOFAR IT (TAB 1, SEZ 5).

Tale investimento andrebbe valorizzato creando le condizioni ottimali per l'accesso ai dati e per la loro analisi allo scopo di ottimizzare l'impatto scientifico per la nostra comunità. L'analisi dei dati LOFAR richiede procedure computazionali molto complesse e infrastrutture adatte alle dimensioni dei datasets prodotti dalle singole osservazioni, tipicamente 10-20 TB. Come dimostrato in questo studio, al momento INAF non possiede/non ha accesso a tali infrastrutture. Per questo motivo abbiamo valutato anche misure sostenibili da affiancare all'investimento di cui sopra per finanziare delle risorse di calcolo di prossimità per l'analisi dati LOFAR che soddisfi le richieste di 10-20 staff INAF e loro gruppi. In questo caso il costo **totale dell'operazione LOFAR per INAF è valutabile fra i 2.55 e i 2.95 MEuro in 5 anni**, a seconda che si preveda anche un investimento per il personale scientifico e al netto di fondi di co-finanziamento esterno già esistenti e del contributo proveniente dal partner UniTo (TAB 2, SEZ 5). Tale costo potrebbe essere ridotto alla luce di un allargamento del consorzio LOFAR IT ad altre Università e di una disponibilità di fondi aggiuntivi di co-finanziamento esterno.

# INDICE

## 1. INTRODUZIONE

- 1.1 SVILUPPI NELLA RADIOASTRONOMIA A BASSA FREQUENZA
- 1.2 IL LOW FREQUENCY ARRAY LOFAR
- 1.3 ITALIAN ROADMAP VERSO LOFAR

## 2. INSTALLAZIONE DI UNA STAZIONE LOFAR 2.0 A MEDICINA

## 3. OTTIMIZZAZIONE DELL'IMPATTO SCIENTIFICO E TECNOLOGICO

- 3.1 COLLABORAZIONE TECNOLOGICA CON ASTRON
- 3.2 OTTIMIZZAZIONE SCIENTIFICA
  - 3.2.1 COINVOLGIMENTO NEI KP
  - 3.2.2 INFRASTRUTTURE PER L'ANALISI DATI LOFAR
  - 3.2.3 SUPPORTO ALLA FORMAZIONE E RICERCA

## 4. CONSORZIO LOFAR IT

## 5. STIMA DEI COSTI

- 5.1 STRATEGIE PER LA RIDUZIONE DEI COSTI
- 5.2 CONCLUSIONE SUI COSTI

## APPENDICI

- 1. INFRASTRUTTURA DI CALCOLO E RETE
- 2. COINVOLGIMENTO TECNOLOGICO IN LOFAR 2.0 & STAZIONE A MEDICINA

## ALLEGATI

- 1. WHITE PAPER 2008
- 2. LoI PREMIALI 2016
- 3. AGREEMENT INAF-ILT
- 4. AGREEMENT INAF-AstroTec
- 5. AGREEMENT INAF-ASTRON

# 1. INTRODUZIONE

L'Italia è uno dei principali sponsor di SKA, una scelta strategica molto importante che è pensata per permettere alla nostra comunità, INAF in primis, di far parte di un grande progetto che permetterà di fare scienza e tecnologia di altissimo impatto nelle prossime decadi. Di fatto SKA ha dato una spinta molto forte alla radioastronomia generando un insieme di precursori/pathfinders che oggi e per almeno una decade guideranno la ricerca di frontiera in ambito radioastronomico. Il coinvolgimento in queste grandi infrastrutture permette alle comunità che ne fanno parte di fare scienza di altissimo livello e di affrontare grandi sfide tecnologiche, ma anche di allevare una generazione di ricercatori che sia in grado di affrontare problematiche della radioastronomia moderna massimizzando l'impatto futuro di SKA sulle comunità.

Al momento l'Italia è l'unico Paese fondatore/finanziatore di SKA che non partecipa a precursori/pathfinders. Questa è una grave anomalia che rischia di precludere alla comunità astronomica italiana una partecipazione con il dovuto peso a quella moltitudine di sviluppi scientifici e innovazioni tecniche che sono in continua evoluzione e che stanno preparando la strada all'esplorazione scientifica con SKA. Questa anomalia risulta ancor più evidente se si valuta il contributo importante di idee apportato nell'ultimo decennio dalla nostra comunità ai casi scientifici di SKA (FIG. 1).

Tale anomalia va risolta in modo rapido e netto per garantire alla nostra comunità di fare scienza di altissimo livello nella prossima decade sulle tematiche dell'era SKA, mantenendo un ruolo importante all'interno di SKA stesso.

**Il ritardo accumulato a questo proposito dall'INAF impone una partecipazione a precursori e/o pathfinder SKA che siano ancora in fase di evoluzione, ma che allo stesso tempo siano già science ready, ossia in grado di produrre da subito scienza di frontiera.**

Tale ritardo esclude infatti la possibilità di investire in un pathfinder italiano in grado di produrre un impatto scientifico significativo, il cui sviluppo richiederebbe un grande investimento finanziario e di manpower e soprattutto lunghi tempi di realizzazione.

In data 27 Gennaio 2017 il Presidente dell'INAF ha dato mandato al Direttore Scientifico e al Direttore Generale di valutare la possibilità di finanziare la partecipazione italiana a precursori/pathfinder SKA science ready con fondi provenienti dall'Astronomia Industriale (SKA-CTA). In data 24/03/2017 il Direttore Scientifico e il Direttore Generale hanno espresso parere positivo a tale operazione identificando un budget potenziale di circa 2 MEuro.

## 1.1 SVILUPPI NELLA RADIOASTRONOMIA A BASSA FREQUENZA

Nell'ultima decade la maturazione della tecnologia di beam forming applicata agli array di apertura e combinata con la crescita delle capacità computazionali hanno prodotto una evoluzione senza precedenti della radioastronomia alle basse frequenze. Questo ha portato

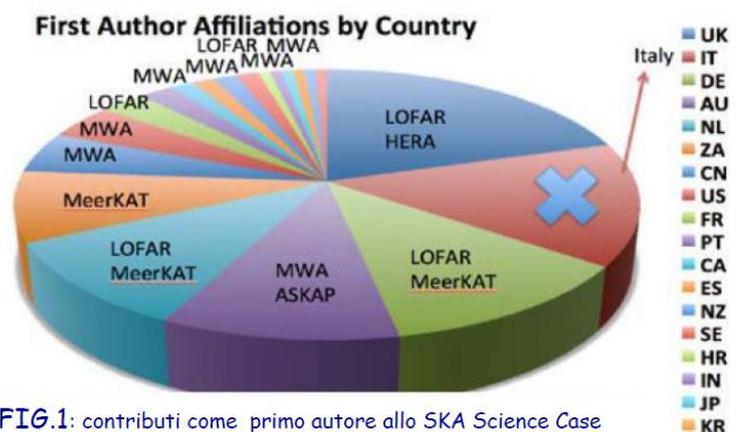


FIG.1: contributi come primo autore allo SKA Science Case raggruppati per Paese con indicata la partecipazione a precursori/pathfinders di SKA

ad un enorme investimento in ambito internazionale con la costruzione di radiotelescopi di nuova generazione (o nell'upgrade di telescopi esistenti) a basse-bassissime frequenze radio mirati ad esperimenti specifici (es il caso della misura di segnali dall'Epoca della Reionizzazione, PAPER, MWA, LOFAR, ..HERA) o osservatori ad ampio spettro (MWA, uGMRT, LOFAR).

Le comunità coinvolte in queste infrastrutture hanno avuto/stanno avendo l'opportunità di affrontare problematiche radicalmente nuove per la calibrazione (es effetti della ionosfera), trasporto e analisi dei dati (es Big Data) che di fatto stanno definendo metodologie totalmente innovative che costituiranno il punto di partenza per risolvere le problematiche che saranno incontrate nell'ambito di SKA-LOW.

Dal punto di vista scientifico queste infrastrutture promettono grandi scoperte se non altro semplicemente perché -di fatto- stanno aprendo una finestra nuova nell'osservazione dell'Universo in un range di frequenze che è di enorme interesse per diverse aree dell'astrofisica e cosmologia, compresi quelli che sono i casi scientifici fondamentali di SKA.

**Non essere coinvolti con il giusto peso in questa straordinaria evoluzione della radioastronomia comporta un grave deficit per la nostra comunità anche in vista di SKA.**

Va anche detto che INAF è molto attiva a livello tecnologico negli array di apertura e beam forming a bassa frequenza (SKADS in passato, SAD, LFAA) ed è coinvolta a livelli importanti nello sviluppo delle antenne e della catena di ricevitori di SKA-LOW (es EMBRACE in passato, AADC-RX). Un coinvolgimento scientifico e tecnologico in uno dei grandi precursori/pathfinder di SKA-LOW darebbe quindi una grande opportunità alla nostra comunità anche per rafforzare la forte sinergia fra tecnologi e ricercatori.

## 1.2 IL LOW Frequency ARray LOFAR

**LOFAR è attualmente lo strumento di punta della radioastronomia a bassa frequenza e il più grande pathfinder di SKA in termini di Area Efficace e di data rate prodotto.**

È un array di apertura completamente digitale organizzato in Stazioni (FIGURA 2) che sono distribuite su 3 scale spaziali: (1) un Core che si estende su un diametro di alcuni km, (2) circa 40 Stazioni remote che si estendono su territorio Olandese, (3) 13 Stazioni internazionali in Germania, Polonia, UK, Francia, Svezia, Irlanda (una 14ma Stazione è già prevista in Lettonia).

LOFAR ha 2 tipi di antenne (FIG 2), LBA (sensibili fra 10-80 MHz) e HBA (100-250 MHz), con un'area efficace di circa 70000 m<sup>2</sup>; per confronto il suo equivalente nel cielo sud, MWA, ha un'area efficace 20 volte più piccola, di circa 3000 m<sup>2</sup>.

I segnali delle antenne LBA e HBA vengono correlati all'interno di ciascuna Stazione e poi inviati in tempo reale tramite collegamento in fibra veloce ad un correlatore centrale (in Olanda) dove i

segnali vengono combinati. Tutta questa catena genera un flusso di dati che non ha precedenti nel campo delle infrastrutture di tipo astronomico (oltre 1 TB/s di dati scambiati all'interno delle stazioni e fra stazioni) ed un flusso di dati astronomici calibrati (e compressi) da archiviare di circa 2 TB/ora per osservazione, richiedendo una capacità di archiviazione di

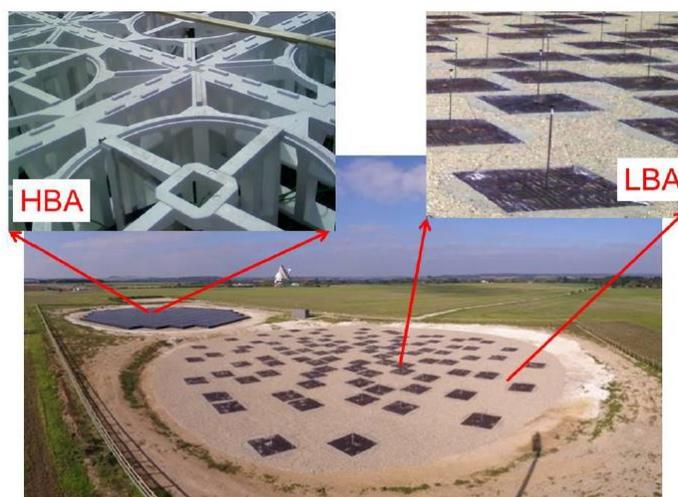


FIG.2: stazione LOFAR Internazionale con particolari delle antenne LBA e HBA

almeno un ordine di grandezza superiore a qualunque altro osservatorio astronomico esistente. Questo rende LOFAR un pathfinder di SKA attualmente unico anche nell'ambito del trasporto e dell'archiviazione dei dati (Big Data problems).

I dati LOFAR sono attualmente archiviati in 3 centri in Olanda (SARA), Germania (JULICH) e Polonia (POZNAN). FIG 3 mostra l'andamento temporale della quantità di dati astronomici archiviati da LOFAR, l'andamento riflette un lungo periodo di commissioning, fino all'inizio del 2016, con un tasso di dati accumulati relativamente basso (considerando le capacità strumentali), ed un grosso incremento negli ultimi anni che dimostra un andamento "a regime" dell'efficienza osservativa del radiotelescopio.

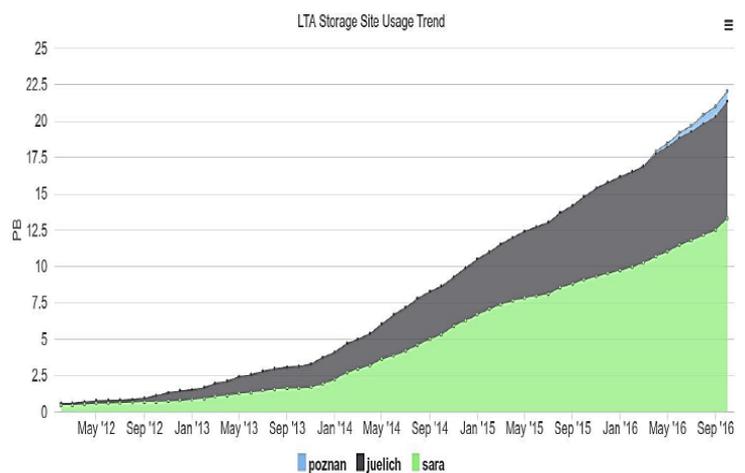
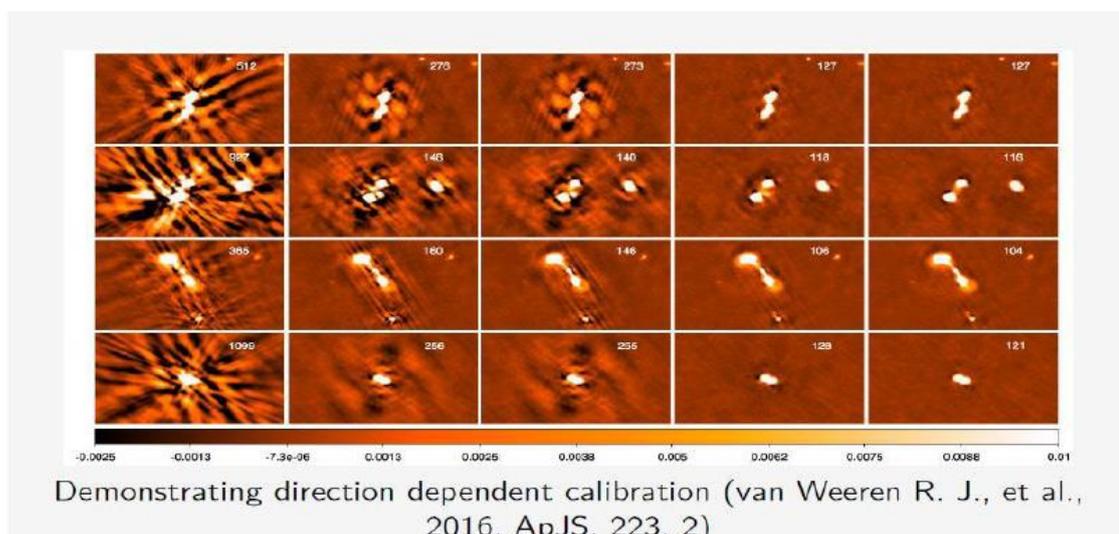


FIG.3: andamento col tempo dei PB archiviati nei 3 archivi ILT presso SARA (verde), JUELICH (grigio) e POZNAN (azzurro).

LOFAR è l'unico radiotelescopio esistente con la capacità di produrre immagini a bassa frequenza con risoluzione angolare di alcuni secondi d'arco (inferiore al secondo d'arco con l'utilizzo delle Stazioni internazionali). Questo rende estremamente complesso il processo di analisi dei dati perché richiede una rimozione capillare delle interferenze (a queste frequenze prodotte da sorgenti anche a grandissime distanza dalle Stazioni riceventi) e una correzione delle distorsioni di fase prodotte dalla ionosfera. Al momento esistono delle complesse catene di pipelines per l'analisi dei dati che sfruttano nuovi concetti sviluppati per la calibrazione delle sorgenti e che permettono di ottenere risultati fino a pochissimi anni fa impensabili permettendo di sfruttare appieno le potenzialità del telescopio. Un esempio è riportato in FIG 4 che mostra l'evoluzione delle immagini di radiosorgenti a 140 MHz (e dell'rms delle immagini riportato nei riquadri in  $\mu\text{Jy}/\text{beam}$ ) dopo l'applicazione di diversi cicli di *direction dependent calibration (facet calibration)*. Il prezzo da pagare al momento per queste procedure è quello di un grande dispendio computazionale la cui ottimizzazione nei prossimi anni costituisce un passo necessario verso SKA. Queste procedure derivano da uno straordinario know-how che è stato sviluppato principalmente all'interno dei KP e di



Demonstrating direction dependent calibration (van Weeren R. J., et al., 2016, ApJS, 223, 2)

FIG.4: evoluzione dell'immagine di radiosorgenti (da sx a dx) dopo diversi cicli di calibrazione seguendo la tecnica direction-dependent implementata in Facet. L'rms delle immagini è riportato nei diversi pannelli in mJ/beam

## ASTRON e che rappresenta un bagaglio fondamentale in vista di SKA-LOW.

A livello scientifico la comunità LOFAR è organizzata al momento in 6 Key Science Projects (KP) la partecipazione ai quali è ristretta ai consorzi che partecipano all'International Lofar Telescope (ILT) (oltre a pochi altri ricercatori coinvolti ad personam) :

1. **Epoch of Reionisation**: ha come scopo principale quello di cercare i segnali (es nello spettro del fondo cosmico radio) dell'epoca della reionizzazione fra  $z=6$  e 12;
2. **Survey Key Project (SKP)**: ha come scopo quello di produrre delle surveys a diversa profondità e risoluzione spaziale del cielo nord;
3. **Transient sources**: ha lo scopo di cercare sorgenti variabili sia galattiche che extragalattiche;
4. **Ultra high energy cosmic rays**: ha lo scopo di studiare i raggi cosmici di alta e altissima energia studiando l'emissione radio generata da sciami nell'atmosfera che vengono prodotti dall'arrivo dei raggi cosmici;
5. **Solar science & space weather**: ha come scopo lo studio del sole, dello space weather e della ionosfera terrestre tramite osservazioni di bassa frequenza;
6. **Cosmic magnetism**: ha lo scopo di studiare le proprietà dei campi magnetici nelle sorgenti galattiche ed extragalattiche tramite osservazioni in polarizzazione a bassa frequenza.

Il SKP è il maggiore dei KP di LOFAR ed è anche quello che in prospettiva richiamerebbe il maggiore interesse da parte della nostra comunità in termini di numero di ricercatori potenzialmente interessati, per questo motivo sarà usato come riferimento scientifico e computazionale in questa relazione.

La FIG 5 mostra un confronto in termini di sensibilità e risoluzione angolare fra la survey Tier-1 del SKP (LOFAR T1)

e quelle esistenti (o pianificate nei prossimi anni, ASKAP/EMU e Apertif) a diverse frequenze; la dimensione dei pallini rappresenta la risoluzione angolare delle surveys. Le survey di LOFAR prevedono una copertura di tutto il cielo nord combinando puntamenti di circa 8 ore di posa (Tier-1), e puntamenti di follow-up più profondi (Tier-2 e 3) di aree più ristrette di cielo tipicamente coperte da follow-up ottici, infrarossi e X (es Bootes, ATLAS-Herschel, Lockman Hole etc). La survey Tier-1 ha una sensibilità e risoluzione angolare 30-100 volte migliori delle surveys preesistenti alle basse frequenze radio (circa 0.1 mJy/beam con 5 arcsec di risoluzione) e paragonabili alle surveys future con i precursori SKA a più alta frequenza (più profonda per sorgenti a spettro ripido,  $\alpha > 1$ ). La survey Tier-1 è in fase di realizzazione e richiederà alcuni anni per essere ultimata. La FIG 6 mostra un confronto fra la distribuzione dei puntamenti previsti per la survey Tier-1 del SKP (punti neri) e i circa 400 puntamenti già effettuati (punti rossi, di cui solo il 10 per cento circa è stato analizzato); in

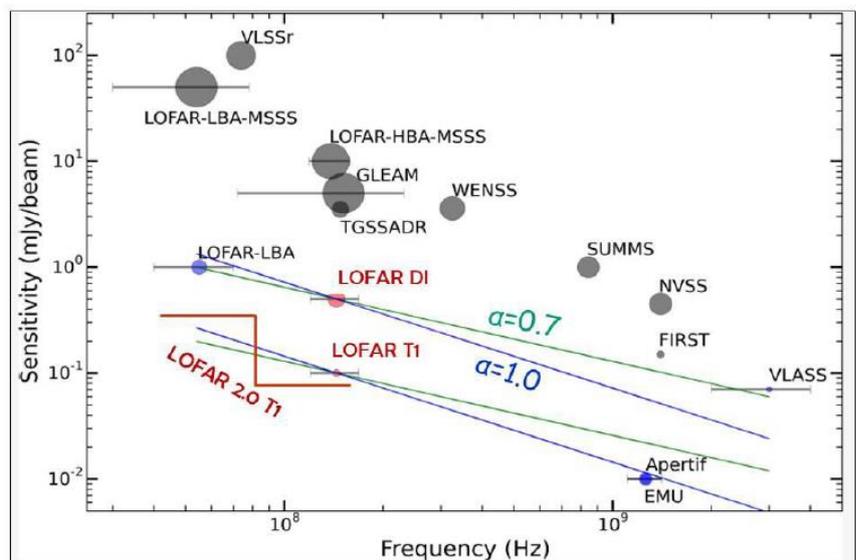


FIG.5: confronto fra la sensibilità delle surveys esistenti a diverse frequenze (include le future Apertif, EMU e VLASS) con le surveys di LOFAR (in particolare la Tier-1). I pallini indicano la risoluzione angolare delle surveys. La sensibilità raggiungibile con LOFAR 2.0 (in una configurazione Tier-1) è indicata in rosso.

FIG 6 è anche indicata la regione (zona blu) coperta dal follow-up spettroscopico che sarà effettuato nei prossimi anni dal progetto WEAVE-LOFAR con il telescopio Herschel da 4.2 mt. **Dalle survey LOFAR ci si aspetta un impatto fondamentale in moltissime aree di ricerca; le seguenti aree sono identificate al momento come prioritarie nell'ambito del SKP e sono organizzate in working-groups :**

- High redshift radio sources;
- Cluster-scale radio emission;
- Starforming galaxies at moderate and high redshift;
- AGN at moderate redshift;
- Detailed studies of AGN at low redshift;
- Nearby galaxies;
- Gravitational lensing;
- Galactic radio sources;
- Cosmological studies from radio surveys.

**Molte di queste aree coincidano con linee di ricerca prioritarie anche in ambito INAF.**

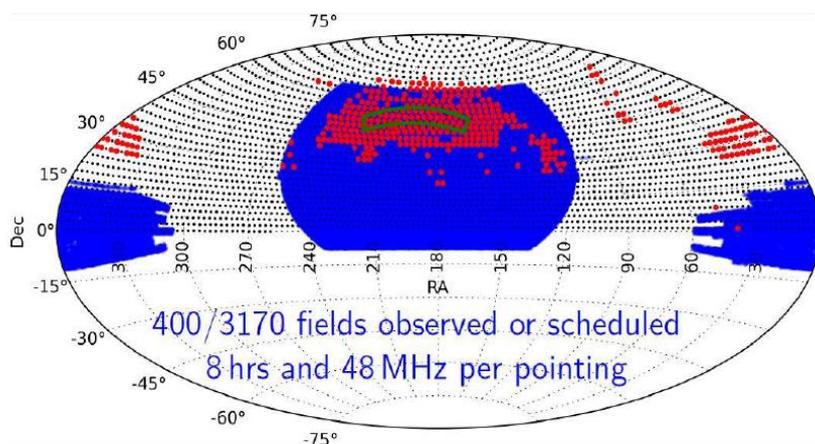


FIG.6: puntamenti pianificati per la survey Tier-1 (in nero) e stato di avanzamento della survey (in rosso). Le zone blu sono le aree di cielo che saranno coperte dal follow-up ottico con WEAVE-LOFAR.

L'infrastruttura di ILT è in continua evoluzione, solo alcuni mesi fa è stata inaugurata e resa operativa la Stazione in Irlanda e nel corso del 2018 sarà ultimata e resa operativa una Stazione in Lettonia.

Da circa un anno il Board di ILT ha deciso un significativo upgrade dell'infrastruttura prevedendo un potenziamento importante della catena di ricezione, delle componenti elettroniche, delle modalità osservative e dell'infrastruttura della correlazione dati. Questo è il primo upgrade sostanziale di LOFAR e prende il nome di LOFAR 2.0, un'operazione che porterà ad un aumento della sensibilità dello strumento (FIG 5), soprattutto nella banda a più bassa frequenza fra 10-80 MHz, e che dovrebbe diventare operativo a partire dal 2021-22.

Fin dal 2007-08, quando LOFAR era ancora in fase di costruzione in Olanda, una massa critica significativa all'interno della comunità Italiana ha mostrato grande interesse verso le potenzialità scientifiche dello strumento, come dimostrato da diverse iniziative intraprese in passato (come esempio il white paper italiano finalizzato nel 2008 per conto del WG nominato da INAF con DP n.19/06 del 09.06.06 "LOFAR: INTEREST AND PERSPECTIVES FOR THE NATIONAL COMMUNITY", Allegato 1).

I diversi Consigli Scientifici che si sono succeduti nel corso degli ultimi 10 anni hanno identificato LOFAR come una delle priorità per la radioastronomia Italiana, anche alla luce della sua importanza come pathfinder di SKA. Più di recente, a Luglio del 2016, circa 70 dipendenti e associati INAF (ricercatori e tecnologi a tempo indeterminato) hanno sostenuto una lettera di intenti per presentare un progetto premiale per acquisire una stazione LOFAR in Italia (unico progetto radioastronomico presentato, Allegato 2).

**In conclusione riteniamo che per le potenzialità strumentali e impatto scientifico atteso in aree di ricerca fondamentali in ambito INAF, per l'importanza come precursore tecnologico e scientifico di SKA-LOW e per la sua natura Europea, è evidente che LOFAR rappresenta la scelta naturale per INAF per partecipare ad un precursore SKA science ready.**

### 1.3 ITALIAN ROADMAP VERSO LOFAR

A partire da Giugno 2017 il Direttore Scientifico dell'INAF ha dato inizio ad una fase esplorativa per definire una possibile roadmap per l'ingresso INAF in ILT.

**Il DS ha indicato come obiettivi quelli di procedere al più presto in un ingresso di INAF in LOFAR e nei suoi KPs installando una o più stazioni LOFAR in Italia.**

La partecipazione di una comunità al consorzio ILT avviene attraverso le seguenti azioni previste dal regolamento dell'ILT:

1. E' necessario costituire un Consorzio Nazionale LOFAR;
2. E' necessario acquisire e installare una o più stazioni LOFAR su territorio Nazionale, garantendo l'operatività di tali strutture e il loro collegamento al correlatore di ILT (Cobalt, ASTRON) tramite reti veloci (alcuni Gbps). Alternativamente è necessario contribuire ad ILT con un investimento (es infrastruttura di calcolo e/o archivio) dal costo equivalente ad una Stazione LOFAR e che sia giudicato congruo e particolarmente utile da parte del BOARD-ILT;
3. I Consorzi Nazionali sono tenuti a pagare una quota annuale (attualmente 92 kEuro per Stazione) per contribuire al funzionamento generale di ILT.

A causa dell'upgrade di LOFAR in LOFAR 2.0 al momento (e per i prossimi 3-4 anni) non è più possibile acquisire una Stazione internazionale. Si è quindi aperta una fase di discussione per capire come ottemperare al punto 2 di cui sopra. Tale fase è continuata anche dopo l'estate 2017, dopo la ristrutturazione dell'UTG 2 di cui la Dr.ssa Govoni è stata nominata coordinatrice e il Dr. Brunetti è stato nominato Advisor relativo ai Precursori e Pathfinder SKA di bassa frequenza. La fase esplorativa si è conclusa a fine Settembre 2017 con una proposta di roadmap concordata fra il Direttore di ILT e l'UTG2 e preventivamente vagliata dal Board di ILT.

Tale proposta (Allegato 3) prevede :

- I. La costituzione di un Consorzio Nazionale LOFAR IT a guida INAF;
- II. Stabilire un accordo di collaborazione tecnologico 2018-2021 fra INAF e ASTRON per lo sviluppo di LOFAR 2.0 prevedendo uno sforzo da parte di INAF su 3 anni equivalente a 400 kEuro (FTE, costi etc);
- III. Stabilire un accordo programmatico fra INAF e AstroTec per l'acquisto e l'installazione di una Stazione LOFAR 2.0 nel 2021-2022, fissando un cost-cup e opportune clausole di garanzia e salvaguardia delle parti;
- IV. Il pagamento della quota di partecipazione a ILT dal momento dell'ingresso del Consorzio LOFAR IT in ILT.

**Consideriamo la roadmap pienamente confacente agli obiettivi identificati dal DS e molto conveniente per i seguenti motivi :**

- a. Permetterebbe un ingresso immediato di INAF in ILT come Full Member, pur non avendo già da gestire e mantenere una Stazione, acquisendo da subito i diritti equivalenti ai Consorzi con 1 Stazione internazionale LOFAR. Ciò risponderebbe alla necessità urgente di avere accesso ad un precursore SKA science ready. **Al momento il target temporale per l'ingresso in ILT è il 16 Aprile 2018 in concomitanza col Board face-to-face di ILT.**
- b. Permetterebbe da subito l'acquisizione di tempo garantito LOFAR (32 ore a semestre per Stazione LOFAR non vincolati oltre a 14 ore/semestre per supporto dei Long Programs, vedi schema in Allegato 3) e il coinvolgimento della comunità interessata nei KP LOFAR a diversi livelli. Il coinvolgimento nei KP è un passo fondamentale perché è il modo più efficiente per confrontarsi con problematiche scientifiche di frontiera, e con strategie innovative di trasferimento e analisi dati. Questo costituirebbe anche un volano straordinario per la formazione e il mantenimento di una comunità tecnicamente

e scientificamente competitiva proiettata verso l'era SKA. A tal proposito va anche fatto notare che un coinvolgimento di alto livello di scienziati Italiani nei KP di uno strumento di frontiera quale LOFAR aumenterebbe anche le chances di ottenere finanziamenti esterni (es ERC Grants etc).

- c. Il contributo tecnologico a LOFAR 2.0 garantirebbe lo status di Full Member ad INAF in mancanza di una Stazione operativa; tale contributo è stato valutato importante dal Board ILT ed "equivalente" a quello apportato da una Stazione funzionante. Come da Allegato 3 il contributo è "one-time entry fee" e in combinazione con l'impegno di acquisto (Allegato 4) garantirebbe lo status di Full Member anche in caso di ritardi nella consegna della Stazione LOFAR 2.0 da part edi AstroTec.

Dalla prospettiva di INAF un investimento tecnologico è molto conveniente perché permetterebbe di rivestire un ruolo centrale anche tecnologico all'interno di LOFAR con conseguenze importanti in termini di know-how e peso politico della comunità tecnologica INAF. Notiamo -infatti- che in passato l'assenza di un coinvolgimento tecnologico è sempre stato considerato un punto a sfavore della partecipazione di INAF in LOFAR.

A valle della conclusione positiva dei negoziati esplorativi, in data 20 Ottobre 2017 la Coordinatrice della UTG2 ha nominato un WG coordinato dal Dr. Brunetti e costituito dai seguenti ricercatori/tecnologi INAF: Dr.ssa A.Bonafede (IRA), Dr. U.Becciani (OA Catania), Dr. P.Bolli (OA Arcetri) Dr. J.Monari (IRA), Dr. M.Nanni (IRA), Dr. F. Perini (IRA), Dr. G.Taffoni (OA Trieste). Lo scopo del gruppo è stato quello di studiare uno scenario tecnologico e scientifico-organizzativo per ottimizzare la partecipazione di INAF in ILT nell'ambito della roadmap di cui sopra. In particolare il WG si è concentrato su 3 punti fondamentali che costituiscono parte integrante di questa relazione :

- Obiettivi, modalità, benefici e sostenibilità della collaborazione tecnologica con ASTRON su LOFAR 2.0. **Questo ha anche portato alla definizione di una bozza di accordo con ASTRON che è stato già vagliato dalla UTG2 e DS;**
- Analisi dati LOFAR in Italia, organizzazione e investimenti necessari. **Questo ha portato a proporre un modello di organizzazione Nazionale, anche a valle del coinvolgimento dell'UniTo in LOFAR IT, e ad una valutazione dei costi;**
- Installazione, connessione alla rete ILT e operatività di una Stazione LOFAR 2.0 a Medicina. **Questo ha anche contribuito alla definizione di una bozza di accordo con AstroTec per l'acquisizione di una Stazione LOFAR 2.0 nel 2021-2022.**

I dettagli degli studi del WG sono riportati in Appendice 1 e 2 .

## 2. STAZIONE LOFAR 2.0 A MEDICINA

L'installazione di una (o più) Stazione LOFAR è uno degli elementi essenziali che sono richiesti per afferire all'ILT (Sezione 1.3).

Come discusso in Sezione 1.3 la roadmap accordata con ILT prevede l'acquisizione di una Stazione di tipo LOFAR 2.0 da installare in Italia possibilmente entro il 2021.

Dopo una serie di negoziati atti anche a creare le condizioni per ridurre possibili fonti di rischio, il cost-cup attuale della Stazione è stabilito in **circa 1.65 MEuro** (Sez.5). Questa stima è stata valutata da AstroTec conteggiando diversi rischi da qui a 4 anni e di conseguenza il costo reale sarà ragionevolmente più basso, probabilmente dell'ordine di 1.4-1.5 MEuro. La bozza di accordo fra INAF e AstroTec con i dettagli è allegata alla relazione (Allegato 4).



**FIG.7:** posizione delle attuali stazioni LOFAR (in arancione le stazioni Olandesi, in blu le stazioni Internazionali) con indicate le principali baselines ottenibili con una stazione Internazionale a Medicina. I 2 cerchi esterni indicano le massime baseline EW ottenibili con l'attuale configurazione di LOFAR.

Il luogo dove installare una Stazione LOFAR deve avere delle caratteristiche ben precise :

- **Baseline:** è strategico posizionare la Stazione in un punto che ottimizzi la copertura uv dell'ILT, ad esempio LOFAR soffre della mancanza di baseline NS su scala di 1000 km (Figura 7);
- **Terreno:** una Stazione LOFAR necessita di circa due ettari di terreno livellato e facilmente accessibile anche da grandi mezzi (ad es per il trasporto delle antenne etc), in una zona radio protetta (protetta da RFI) e servita da rete per energia elettrica;
- **Struttura e personale:** preferibilmente una Stazione LOFAR dovrebbe essere posizionata in una zona facilmente raggiungibile da tecnici, o dove la presenza di tecnici specializzati (in regime di protezione radio) e' già assicurata da insediamenti precedenti (ad es in vicinanza di stazioni radioastronomiche o per telecomunicazioni già operative);
- **Fibra ad alta velocità:** è necessario connettere una Stazione LOFAR ad una rete ad alta velocità, tipicamente 10 Gbps di cui una banda di almeno 3 Gbps dedicati interamente alla Stazione, collegata con le principali dorsali Italiane e Europee per il collegamento con le infrastrutture di calibrazione in Olanda.

Sulla base di questi punti la migliore scelta è data da Medicina (BO) dove sono già ospitate le infrastrutture radioastronomiche (in regime di protezione radio) Croce del Nord e l'antenna VLBI da 32 mt dell'IRA-INAF (Figura 8). Per completezza menzioniamo che la stessa conclusione è stata ottenuta alcuni anni fa da un WG INAF indipendente (DP n.19/06 del 09.06.06, Chair Dr. R. Scaramella).



Nello specifico :

- **Baseline:** come mostrato in FIG 7, la posizione di Medicina permette una buona espansione della copertura del piano uv di LOFAR, con una baseline che espande la configurazione dello strumento in direzione N-S, ma che è paragonabile all'estensione delle baseline in direzione E-W e che quindi non creerebbe eccessivi gap nella copertura del piano uv di LOFAR. Una volta installata la Stazione a Medicina, ulteriori Stazioni potrebbero essere installate in Piemonte (per ottimizzare la copertura della baseline intermedia fra Medicina, Nancay e Garching) o in Sardegna o Sicilia (soprattutto se la Spagna dovesse aderire ad ILT installando una o più Stazioni LOFAR 2.0).
- **Terreno:** Medicina offre una soluzione ottimale essendo facilmente ricavabile uno spazio dell'ordine di 1-2 ettari per ospitare la Stazione nei dintorni delle attuali infrastrutture radio. E' stata individuata un'area (vedi FIG 8) dove posizionare una Stazione LOFAR 2.0 che per l'assenza di vegetazione di alto fusto e per la presenza di un accesso minimizzerebbe anche i costi di preparazione del terreno stimabili in circa 60 kEuro (Sez.5).
- **Struttura e personale:** alla Stazione radioastronomica di Medicina afferisce la maggior parte del personale del gruppo tecnologico INAF per le antenne e ricevitori a bassa frequenza e quindi lo stesso personale che sarebbe coinvolto nella collaborazione tecnologica su LOFAR 2.0.
- **Connessione dati:** la stazione di Medicina è collegata alla rete GARR con una lambda da 10 Gbps per i dati. Tale connessione dati è tecnicamente adeguata per supportare contemporaneamente i flussi delle osservazioni eVLBI e LOFAR. Il costo per il trasporto dati continuativo sulle reti GARR e GEANT verso l'Olanda, considerando un flusso continuativo di circa 4 Gbps, è di circa 20 kEuro/anno (Appendice 2).

**FIG.8:** vista aerea della Stazione radioastronomica di Medicina. Si vedono la Croce del Nord con gli edifici che ospitano il personale e l'antenna da 32 mt. Il fotomontaggio mostra la stazione LOFAR.

### 3. OTTIMIZZAZIONE DELL'IMPATTO SCIENTIFICO E TECNOLOGICO

L'ingresso di INAF in ILT sarebbe un investimento molto significativo nell'ambito della radioastronomia. Per questo motivo e per le grandi opportunità che tale investimento offrirebbe alla nostra comunità anche in vista di SKA e dell'utilizzo scientifico di altri precursori/pathfinders, è necessario ottimizzarne l'impatto sia in termini scientifici che tecnologici. A tale scopo, sulla base della roadmap in Sezione 1.3, riteniamo particolarmente strategici:

1. Rivestire un ruolo tecnologico chiaro e di rilievo nello sviluppo di LOFAR 2.0 con un impatto significativo sul know-how della nostra comunità e sulla possibilità di cogliere opportunità per intercettare fondi esterni (es ERC);
2. Formare una massa critica di ricercatori INAF, anche al di fuori delle classiche sedi di radioastronomia, in grado di analizzare e utilizzare i dati LOFAR in molteplici ambiti;
3. Usare l'esperienza di LOFAR per sviluppare know-how nella nostra comunità nel campo degli algoritmi di analisi di Big Data e del loro trasferimento e archiviazione.

Riteniamo che per soddisfare questi punti siano necessarie 4 azioni specifiche :

- i. Definire un accordo di collaborazione tecnologica con ASTRON su aspetti importanti di LOFAR 2.0 che preveda ruoli di PI-ship per INAF e che coinvolga INAF su un ampio spettro della roadmap di LOFAR 2.0;
- ii. Coinvolgere una massa critica di ricercatori INAF negli attuali KP di LOFAR;
- iii. Dotare INAF di una infrastruttura su scala Nazionale adeguata all'analisi e gestione dati LOFAR, costituendo anche una task-force per l'installazione, supporto, gestione e sviluppo delle pipeline di analisi dati;
- iv. Supportare attività specifiche per la formazione di giovani ricercatori; quest'ultimo punto è particolarmente importante per i gruppi non particolarmente attivi nel campo della radioastronomia.

#### 3.1 COLLABORAZIONE TECNOLOGICA CON ASTRON

L'attuale roadmap prevede di garantire la Full Membership di INAF in ILT fin da subito tramite il coinvolgimento tecnologico in LOFAR 2.0 (punto II, Sez 1.3), prevedendo un investimento di circa 400 kEuro.

Da Ottobre 2017 sono stati fatti numerosi incontri (face-to-face e telecon) per individuare con la controparte ASTRON le possibili linee di ricerca sulle quali indirizzare la collaborazione tecnologica su LOFAR 2.0 che avrà **una durata di circa 3 anni**.

Alla luce di quanto detto prima (punti 1. e i. di cui sopra), riteniamo che un ambito strategico e naturale per INAF per una collaborazione tecnologica con ASTRON su LOFAR 2.0 sia quello dello sviluppo di antenne di bassa frequenza e loro calibrazione, e della catena di ricezione analogico-digitale di LOFAR 2.0. In tale ambito gli ingegneri INAF hanno grande competenza e la collaborazione sarebbe fortemente connessa alle attività in itinere su SKA-LOW e ne sfrutterebbe pienamente le competenze sviluppate in ambito INAF.

Le discussioni sono ancora in itinere ed è naturale che molti dei dettagli potranno essere definiti solo nei prossimi anni. Tuttavia le parti hanno già prodotto un testo di accordo programmatico identificando le linee di collaborazione chiave, le modalità della collaborazione e i prodotti attesi (Allegato 5). Di seguito riassumiamo le linee principali della collaborazione (vedi Appendice 2 per maggiori dettagli):

1. **Catena di Ricezione RCU analogica** : è la principale linea di collaborazione ed inizierebbe subito dopo l'estate del 2018. INAF sarebbe coinvolta nella definizione dei requisiti generali di sistema e sarebbe responsabile del design, simulazione e prototipizzazione della parte analogica della catena di ricezione di LOFAR 2.0. Questa è una delle parti più innovative di LOFAR 2.0 e ha forti aderenze di tipo tecnologico con quanto già fatto da INAF in progetti SKA-oriented (es EMBRACE);
2. **Antenne LBA** : INAF collaborerà allo studio elettromagnetico e della componente elettronica delle antenne LBA per definire gli scenari di upgrade per LOFAR 2.0.

La proprietà intellettuale dei prodotti del progetto sarà condivisa fra ASTRON e INAF.

Al momento si stima che il punto 1. della collaborazione coprirà già una frazione significativa (circa il 60%) dello sforzo complessivo richiesto ad INAF (Appendice 2), di conseguenza nel corso del progetto si valuterà l'ammontare dell'investimento (soprattutto FTE) di INAF sul punto 2 e su eventuali linee aggiuntive. Allo stato attuale va chiarito che la collaborazione sul punto 1 richiede complessivamente circa 1.5-1.7 FTE coperti interamente da personale già in servizio presso la sede di Medicina dell'IRA e circa 120 kEuro per costi di funzionamento e prototipizzazione (Appendice 2). **L'accensione di una collaborazione bilaterale fra INAF e ASTRON porrà la nostra comunità in una posizione di forza all'interno della comunità ILT rendendo INAF -al momento- l'unico Istituto non Olandese coinvolto nello sviluppo tecnologico di LOFAR 2.0.** Questo è importante anche perché aprirà alla possibilità da parte nostra di essere coinvolti con ruoli di co-PI-ship in richieste di finanziamento su fondi Europei; facciamo notare a tal proposito che la comunità ILT ha già ottenuto una lista significativa di progetti finanziati su bandi Europei.

### 3.2 OTTIMIZZAZIONE SCIENTIFICA

Premettiamo che la comunità Italiana ha sempre mostrato interesse nelle tematiche dei KPs di LOFAR che peraltro costituiscono vere e proprie attività precursori verso SKA. L'interesse da parte di INAF verso una infrastruttura osservativa con grande potenziale e in grado di aprire una nuova finestra nell'osservazione dell'Universo è peraltro cosa molto naturale.

Dati i vincoli temporali e lo scopo di questa relazione riteniamo inutile ripetere l'esercizio di una ricognizione capillare degli interessi a livello Nazionale alla luce di quanto già discusso in Sezione 1.2, dell'ampia documentazione già raccolta negli ultimi anni (es Allegato 1), delle azioni recenti tentate dalla comunità per accedere a finanziamenti per installare Stazioni LOFAR in Italia (es Allegato 2) e delle valutazioni dei vari Consigli Scientifici dell'INAF che si sono succeduti nell'ultimo decennio.

Discorso meno ovvio è invece quello di suggerire come ottimizzare il ritorno scientifico dell'ingresso della nostra comunità in LOFAR. Come già detto in precedenza questo è un punto cruciale nell'ambito di un modello di roadmap e pertanto è un punto centrale di questa relazione.

**Pertanto scopo di questo report è quello di proporre una roadmap SOSTENIBILE per INAF che fornisca al personale le condizioni NECESSARIE al coinvolgimento scientifico ottimale in LOFAR, garantendo a TUTTI i gruppi interessati di accedere:**

- ❖ ai dati LOFAR e alle pipelines necessarie alla loro riduzione e analisi, principalmente tramite il coinvolgimento nei KPs;
- ❖ ad una infrastruttura Nazionale adeguata all'analisi e archivio dei dati;
- ❖ ad un gruppo Nazionale di supporto per l'analisi dati;
- ❖ a fondi necessari per la formazione di giovani ricercatori appartenenti al gruppo.

**In altre parole la roadmap ha come scopo principale di OFFRIRE ai gruppi INAF interessati alle aree di ricerca di LOFAR e SKA-LOW le condizioni ottimali per**

l'acquisizione e l'analisi dei dati LOFAR e per la formazione dei giovani ricercatori, **RICHIEDENDO** tuttavia uno sforzo di investimento da parte dei gruppi stessi (es in termini di personale se necessario).

E' evidente che un MINIMO finanziamento del personale scientifico e tecnologico sulle tematiche del progetto è quantomeno auspicabile da parte di INAF (es Sez.3.2.3, e Sez.5 TAB 2), tuttavia -date le attuali condizioni al contorno- nuovo personale andrà finanziato principalmente tramite bandi competitivi dell'ente (es PRIN INAF) o fondi esterni.

### 3.2.1 COINVOLGIMENTO NEI KP

I 6 KP di LOFAR (Sezione 1.2) sono il riferimento principale dell'investimento in LOFAR. Essi costituiscono attualmente le comunità dove esistono le maggiori competenze tecniche e dove si affrontano le problematiche in modo innovativo sviluppando un importante know-how anche in vista di SKA. I KP sono comunità in continua evoluzione dove le competenze dei vari gruppi vengono valorizzate acquisendo peso e ruoli di PI-ship di linee di ricerca. **E' importante quindi il coinvolgimento di una massa critica di ricercatori INAF nei KP di LOFAR per sviluppare competenze e per apportare un contributo anche importante ai KP stessi.** *Questo coinvolgimento diventerebbe automaticamente possibile a valle dell'ingresso di INAF in ILT.* Le modalità di reclutamento dei ricercatori INAF nei KP andranno definite in ambito Nazionale e in accordo con i PI dei vari KP, eventualmente definendo anche linee di ricerca aggiuntive a guida Italiana (oppure co-PI-ship di linee di ricerca esistenti). Premesso che un coinvolgimento troppo esteso e indiscriminato sarebbe controproducente, una possibile strada sarebbe quella di una *Call for Interest Nazionale* gestita dall'Ufficio della Direzione Scientifica (o dalla UTG2) dalla quale sia possibile valutare il contributo proposto da ciascun ricercatore in termini di manpower, idee e competenze per procedere ad un ranking.

Al momento 6 ricercatori INAF (TI e TD presso IRA, OACagliari, OARoma) sono coinvolti in KPs di LOFAR (Surveys e Magnetismo) con ruoli molto diversi e a diversi livelli. La partecipazione "da esterni" non ha comunque consentito di creare una massa critica in Italia per l'utilizzo di LOFAR ed infatti esperienze di analisi di dati LOFAR sono state fatte solo recentemente presso l'IRA. Tuttavia **tali esperienze e le fortissime connessioni già presenti con i KPs da parte di alcuni di questi ricercatori consentirebbero un rapido sviluppo delle capacità di analisi dati in Italia e un coinvolgimento della comunità molto ampio alla luce dell'ingresso di INAF in ILT e della presenza di infrastrutture computazionali adeguate su scala Nazionale** (Sez. 3.2.2).

Riteniamo che l'obiettivo **MINIMO** commisurato all'investimento sia quello di coinvolgere in modo attivo nei KP circa 10-12 ricercatori TI e i loro studenti e post doc, mentre il **coinvolgimento di una 20na di ricercatori rappresenterebbe un OTTIMO risultato.** L'interesse dei ricercatori INAF si concentra principalmente nel Survey KP (vedi Sez 1.2) e in misura meno significativa nel Transient KP e nel Magnetism KP. Inoltre, una buona parte dei nuovi ricercatori coinvolti nei KP si attiverà principalmente in attività di follow-up delle sorgenti ad altre frequenze (es ottico, IR, high-energy).

Date queste condizioni al contorno, nelle prossime Sezioni per costruire un modello funzionale assumeremo che circa 20-25 ricercatori INAF, compresi staff, post-doc e PhD siano attivi nell'analisi di dati LOFAR. Questa sarebbe un'ottima massa critica, superiore anche a quella di molti altri Paesi attualmente coinvolti in LOFAR.

### 3.2.2 INFRASTRUTTURE PER L'ANALISI DATI LOFAR

L'ingresso di INAF in ILT porterebbe a 32 ore di tempo garantito per semestre (Allegato 3) che genererebbero circa 120 TB di dati all'anno per i prossimi 5 anni, assumendo come riferimento una modalità di osservazione e acquisizione dati tipica delle Survey Tier-1 di LOFAR HBA. **Il coinvolgimento della nostra comunità nei KP secondo le linee delineate in Sect.3.2.1 porterebbe ad un flusso di dati ancora maggiore, stimabile in circa 200-400 TB di dati all'anno considerando dati da KP e Open Time (OT).**

I dati sarebbero archiviati nelle infrastrutture afferenti a ILT (attualmente i centri di calcolo/archivio di SARA (NL), JULICH (D), e POZNAN (P)). In linea di principio tramite le infrastrutture di calcolo di ASTRON l'ILT garantisce le risorse minime/di sopravvivenza per l'analisi a tutti gli utenti che ottengono dati da OT. Tuttavia è evidente che per supportare l'analisi avanzata di 200-400 TB/anno di dati ad una comunità intera e per permettere a tale comunità di acquisire un ruolo importante nei KP è necessario accedere a strumenti adeguati.

**Come discusso in Appendice 1, al momento INAF non ha accesso ad una infrastruttura in grado di supportare in modo efficiente la riduzione del flusso di dati che è previsto da LOFAR. Per questo motivo - per garantire un ritorno appropriato dall'investimento su LOFAR - riteniamo che sia NECESSARIO dotarsi di un risorse di calcolo di prossimità adeguate.** Di conseguenza la formulazione di un modello sostenibile di infrastruttura per l'analisi dati su scala Nazionale è stato uno degli scopi più importanti di questo studio.

Abbiamo formulato un modello di infrastruttura usando come riferimento le **pipelines di analisi delle osservazioni LOFAR HBA nel continuo adottando la tipica configurazione delle osservazioni del Survey KP (si veda l'Appendice 1 per dettagli)**. In accordo con quanto detto sopra, abbiamo assunto una situazione in cui il flusso di dati LOFAR verso la nostra comunità si attesta su alcuni x100 TB/anno e che vi siano 20-25 utenti attivi nell'analisi dati.

Lo studio delle pipelines ha evidenziato una chiara divisione del processo di analisi dati LOFAR in 2 parti, fase 1 e 2, che possono essere separate chiaramente in termini di

- tipo di interazione fra l'utente e il calcolatore/programma di analisi;
- dimensione iniziale dei dataset;
- quantità di RAM utilizzata durante il processo di analisi dati;

Nel corso della **fase 1** i dati iniziali sono tipicamente dell'ordine di 10-20 TB. La procedura di riduzione dati produce un dataset di circa 500 GB tramite opportuni passi (initial subtraction e pre-factor) durante i quali sono attivate procedure di flag, calibrazione, compressione dati, sottrazione delle sorgenti e sono ottenute prima immagini ad alta e bassa risoluzione e modelli di auto-calibrazione. Questa fase è altamente RAM intensive, prevede pochissime interazioni con l'utente e dura mediamente 5-10 giorni di tempo su un tipico cluster LOFAR (Appendice 1). Nel corso della **fase 2** vengono attivate particolari fasi di calibrazione dei dati in funzione della direzione delle sorgenti (direction-dependen calibration, facet-calibration) per produrre immagini di alta qualità partendo dai prodotti ottenuti in fase 1. La fase 2 è computing intensive (RAM intensive ma meno della fase 1) e prevede un livello di interazione significativa con l'utente. In questa fase possiamo anche includere la vera e propria fase finale di analisi dei dati, inclusa la produzione di molteplici immagini ottenute con diverse pesature delle baseline di LOFAR, e di manipolazione di immagini (misure flussi, spettri, mappe, sovrapposizioni etc); tale fase finale fa uso di packages standard quali ad esempio CASA etc. Nel complesso la fase 2 richiede tipicamente 1-3 mesi di lavoro durante i quali l'utente ha bisogno di un nodo di calcolo possibilmente dedicato e anche di spazio disco significativo per il salvataggio dei dati.

Attualmente le uniche infrastrutture accessibili ad INAF in grado di consentire la fase 1 e 2 sono il nodo FAT attualmente installato presso l'IRA (che può soddisfare 1-2 utenti) e il

cluster HOTCAT installato presso OATrieste (che consentirebbe però di fare solo parte della fase 2 di cui sopra e che tuttavia non ha nodi dedicati). Sulla base di quanto sopra si propone il seguente modello di risorsa di calcolo di prossimità (Appendice 1) :

**Infrastruttura 1** : una unica **media** infrastruttura di calcolo, collegata alle infrastrutture di calcolo e archivi ILT (ASTRON, SARA, JULICH, POZNAN) via rete ad alta capacità (es, GARR da almeno 10 Gbps), con le seguenti caratteristiche :

- Cluster di Calcolo costituito da nodi con disponibilità di RAM di almeno 512 GB fino ad 1TB. 1 o 2 Socket con processore Intel Xeon di ultima generazione o equivalenti e un numero di core pari ad 8-32 a seconda della quantità di RAM (64-16 GB RAM/core). Interconnessione su rete infiniband 56Gbs (o superiore).
- Storage distribuito ad alte prestazioni di almeno 80 TB eventualmente basato su dischi SSD o SAS.
- Un secondo sistema di storage distribuito a basse prestazioni con capacità fino a diversi x100 TB; tale archivio servirebbe a mantenere i dati iniziali per un breve tempo e a mantenere i prodotti avanzati/finali della fase 1 e 2 per medio-lungo tempo.

**Risorse di tipo 2** :

- Clusters di Calcolo costituiti da nodi con disponibilità di RAM di almeno 256 GB. 2 Socket con processore Intel Xeon di ultima generazione o equivalenti e un numero di core pari ad almeno 8-16 per socket (16 GB RAM/core). Interconnessione su rete infiniband 56Gbs (o superiore).
- Storage distribuito ad altissime prestazioni di almeno 10 TB eventualmente basato su dischi SSD o SAS.
- Un secondo sistema di storage distribuito a basse prestazioni di almeno 50 TB.
- Remote desktop.
- Connessione alla rete GARR ad almeno 1 Gbps (auspicabili a 10 Gbps).

### **ORGANIZZAZIONE:**

l'infrastruttura combinata 1+2 permetterebbe di supportare pienamente una 20na di utenti, dando anche la possibilità di archiviare su medio periodo sia i dati provenienti dagli archivi ILT (Infrastruttura 1) sia i prodotti dell'analisi dati degli utenti (Risorse di tipo 2). Per ottimizzare il supporto ad una utenza distribuita su scala Nazionale si suggerisce l'installazione sui nodi di *remote desktops* che hanno un costo contenuto e permettono un uso delle facilities in remoto, dagli Osservatori e/o Istituti INAF, anche in caso di interazione continua utente-macchina.

La infrastruttura 1 dovrebbe essere unica, al contrario le risorse di tipo 2 potrebbero essere pensate come uniche oppure distribuite su più sedi a livello Nazionale.

Una infrastruttura di calcolo di media grandezza e con nodi FAT adeguati alla **fase 1** di analisi è già accessibile - in linea di principio - all'attuale partner del consorzio LOFAR-IT (UniTo, Sez.4) e risulta quindi evidente che **affidare l'Infrastruttura 1 a UniTo sia una scelta efficiente garantendo anche una minimizzazione dei costi.**

**L'organizzazione delle risorse di tipo 2** andrebbe pianificata nell'ambito di una proposta di roadmap dell'Ente molto più ampia, di medio-lungo termine, sulle infrastrutture di calcolo, analisi dati e archivio coinvolgendo diverse sezioni della DS (ICT, IA2, e altre UTG); questo è comunque al di là dello scopo di questo studio. Ci siamo quindi limitati ad una proposta di organizzazione per LOFAR che fosse immediatamente operativa, facilmente sostenibile e di corto-medio termine. Notiamo che la scelta di un modello unico o distribuito per le risorse di tipo 2 non ha un impatto significativo sull'efficienza del processo di analisi dati. **Proponiamo quindi un semplice modello distribuito di risorse di tipo 2 basato su 3 clusters di calcolo costituiti da 6 nodi ciascuno e distribuiti su scala Nazionale.** Il modello distribuito su più

poli è pensato per massimizzare il coinvolgimento effettivo dei gruppi informatici interessati al progetto ma soprattutto per minimizzare i costi sfruttando cofinanziamenti da gruppi di ricerca locali (ad es IRA già prevede di installare fino a 4-5 nodi LOFAR (1 è già operativo) sfruttando gli overheads di progetti Europei o fondi esterni, e al momento un cofinanziamento per nodi LOFAR sarebbe già disponibile anche presso l'Osservatorio di Catania) (Sez.5, TAB.2). Notiamo inoltre che un modello distribuito sul corto-medio termine minimizzerebbe anche la gestione delle risorse perché un piccolo numero di nodi sarebbe facilmente gestibile dai system manager locali nel caso degli Osservatori o Istituti che già gestiscono facilities di calcolo di piccola dimensione. Un modello non distribuito (ie 18 nodi in una struttura unica) sarebbe comunque ugualmente funzionale per LOFAR e rimane comunque un'opzione percorribile dall'Ente ad un costo superiore (Sez.5, TAB.2).

**L'installazione e gestione delle pipelines** LOFAR è un processo troppo oneroso per un utente e per questo si suggerisce la costituzione di una **task force informatica INAF** che si occupi della installazione, aggiornamento ed operatività delle pipelines sui diversi nodi (Infrastruttura 1 e risorse di tipo 2). Riteniamo inoltre che per massimizzare l'impatto di LOFAR in termini di know-how sia utile un coinvolgimento di INAF all'interno dei gruppi internazionali (ASTRON, Groningen, Amburgo, Meudon) che elaborano il software per l'analisi dei dati LOFAR. In tal senso la *task force informatica* INAF potrebbe occuparsi non solo del supporto informatico ma anche dell'ottimizzazione delle pipeline esistenti e dello sviluppo di nuovi algoritmi. Oltre ad aumentare il peso effettivo di INAF in ILT, questo permetterebbe anche di cogliere importanti opportunità di collaborazione e finanziamenti esterni legati a progetti internazionali su Big Data e grandi infrastrutture computazionali.

### 3.2.3. SUPPORTO ALLA FORMAZIONE E RICERCA

**Il supporto alla formazione di giovani ricercatori per l'analisi dati LOFAR costituisce una condizione NECESSARIA per l'ottimizzazione del ritorno scientifico in INAF.**

Nel caso di LOFAR esistono degli ambiti di riduzione dati (ad esempio l'uso delle lunghissime baselines per ottenere immagini ad altissima risoluzione) che di partenza richiedono già elevatissime competenze nel campo delle tecniche radioastronomiche (es VLBI etc); questi sono ambiti che rimarranno ragionevolmente territorio dei soli radioastronomi. Tuttavia negli ambiti più convenzionali e che rivestono buona parte dell'interesse scientifico il processo di analisi dati è affidato all'uso di un complicato insieme di pipelines da parte dell'utente senza richiedere elevatissime competenze di tipo radioastronomico. In linea di principio questo permette di esportare la capacità di analizzare dati anche al di fuori delle classiche sedi con indirizzo radioastronomico in INAF, con un potenziale incremento del ritorno in termini di impatto scientifico. In tal senso dare l'opportunità di crescita anche a gruppi non tradizionalmente radio diventa una priorità dell'intera roadmap LOFAR.

Per ottenere questo si suggerisce di finanziare le seguenti azioni :

1. 4-6 borse di studio Nazionali all'anno per almeno i primi 2 anni dall'ingresso di INAF in ILT a copertura di periodi di ricerca di 3-4 mesi in centri di eccellenza per l'analisi dati LOFAR (es Leiden, Amburgo, ASTRON, etc). *Questo schema è stato già sperimentato in ambito INAF ed è risultato molto efficace;*
2. finanziare borse di dottorato in ambito INAF/Nazionale su tematiche LOFAR e/o SKA-LOW anche in poli non tradizionalmente radio prevedendo eventualmente anche il coinvolgimento di radioastronomi esperti di bassa frequenza come co-supervisors. *Non si tratterebbe necessariamente di finanziare nuove borse di dottorato ma eventualmente di focalizzarne le tematiche nei prossimi anni anche nell'ambito di accordi già finanziati e in itinere con le Università;*

3. un MINIMO di 2-3 postdoc Nazionali all'anno di durata biennale/triennale per almeno i primi 2 anni dall'entrata di INAF in ILT per finanziare giovani ricercatori coinvolti nei KP di LOFAR o per attrarre figure qualificate dall'esterno;
4. scuole di astrofisica su tecniche di analisi di dati radio con un focus specifico su LOFAR.

## 4. CONSORZIO LOFAR-IT

Come spiegato in Sezione 1.3, per l'ingresso in ILT è necessario costituire un consorzio nazionale LOFAR che sia aperto ad Enti di Ricerca e Università. La prassi seguita a livello internazionale è che uno degli Istituti/Università che costituiscono tale consorzio ne diventa il rappresentante legale e viene delegato dal consorzio per i rapporti formali con ILT (es firmare agreement etc). Allo stesso tempo il consorzio si dà una struttura nella forma di un Board con un chairman che rappresenta il consorzio stesso all'interno del Board ILT.

In Italia la situazione è diversa rispetto a quelli di altri paesi poiché l'INAF è l'Istituto Nazionale delegato dal MIUR a condurre ricerca nel campo dell'Astronomia e Astrofisica, e di fatto raccoglie la maggioranza dei ricercatori nel campo. **Di conseguenza è naturale che INAF guidi la costituzione di un consorzio LOFAR-IT e che -almeno in una fase iniziale- ne sia rappresentante legale nei rapporti con ILT.**

Il coinvolgimento di Università o altri Enti di Ricerca in un consorzio LOFAR IT è importante per aumentare la massa critica dei ricercatori coinvolti e il ritorno scientifico e tecnologico per la comunità Nazionale, ma anche per un'ottimizzazione dei costi di INAF potendo sfruttare infrastrutture, fondi e personale eventualmente messi a disposizione da gruppi/dipartimenti universitari.

Fin dall'estate 2017 la Direzione Scientifica e la UTG2 hanno iniziato una fase di discussione con il **Dipartimento di Fisica dell'Università di Torino (UniTo)** per la costituzione di un consorzio LOFAR IT. Gruppi di ricerca dell'UniTo e personale afferente al Consorzio Interuniversitario per la Fisica Spaziale (CIFS) di Torino avevano già partecipato alla Lettera di Intenti per PREMIALI INAF 2016 su LOFAR (Allegato 2) e sono fortemente interessati a LOFAR sia per aspetti scientifici (AGN, radiogalassie, raggi cosmici) che per aspetti tecnologici (antenne a bassa frequenza).

**Il principale contributo da parte di UniTo nei primi anni sarà sulla infrastruttura di calcolo e archivio, in particolare l'Infrastruttura 1.** UniTo gestisce il Centro di Competenza del Calcolo Scientifico (C3S) che è un centro di ricerca interdipartimentale specializzato in sistemi HPC. Il C3S gestisce OCCAM che è una infrastruttura di calcolo e archivio composta da diversi tipi di nodi di calcolo fra cui FAT e LIGHT Nodes con caratteristiche compatibili a quelle delle infrastrutture di calcolo delineate nella Sezione 3.2.2 (e Appendice 1). OCCAM inoltre ha uno spazio di disco (archivio) di circa 1 PB, anche questo compatibile con le esigenze delineate in Sezione 3.2.2. I negoziati sono in una fase avanzata, l'idea attuale è che tramite il C3S UniTo riserverà l'accesso a 4 nodi FAT di OCCAM a LOFAR IT e uno spazio disco crescente fino a circa 500 TB in 3 anni; i fondi per garantire questo investimento sono in gran parte già disponibili ad UniTo.

**La UTG2 ha anche iniziato ad esplorare l'interesse in LOFAR-IT da parte di ulteriori partner universitari per espandere il consorzio nei prossimi mesi.**

## 5. STIMA DEI COSTI

Come primo dato riportiamo la stima dei costi relativi alla partecipazione in LOFAR secondo le condizioni stabilite con ILT (Sezione 1.3).

Cio' comprende :

1. copertura finanziaria per la collaborazione tecnologica con ASTRON per LOFAR 2.0 (Sezione 3.1, App 2); gli FTE necessari sono già disponibili nei gruppi di Medicina e Arcetri (in blu in TAB 1).
2. Acquisto di una stazione LOFAR 2.0, preparazione del terreno, installazione a Medicina e costi operativi della Stazione a partire dal 2021-22 (elettricità, connessione di rete, FTE, Sezione 2, App 2); gli FTE per l'operatività della stazione sono già disponibili a Medicina (in blu in TAB 1).
3. Running costs per la partecipazione al consorzio ILT, attualmente 92 kEuro/anno.

Va detto che l'attuale stima del costo della Stazione di tipo LOFAR 2.0 è un upper-cup, che lo schema di pagamenti sarà ragionevolmente spalmato fra 2021 e 2022 e che esisterà una flessibilità fino ad 1 anno per la consegna dell'infrastruttura a Medicina (Allegato 4).

TABELLA 1: PREVISIONE DI COSTI PER I PRIMI 5 ANNI

	2018	2019	2020	2021	2022	TOT
TECNO FTE	20 30	90 90	40 90	10 30		160 240
STAZIONE TERRENO RUNNING FTE				1650 60 30	40 15	1650 60 40 45
RUNNING	65	92	92	92	92	453
INAF INVEST	115	272	222	1862	147	2638
<b>INAF COST</b>	85	182	132	1802	132	<b>2353</b>

**Il costo complessivo è di circa 2.3-2.4 MEuro in 5 anni, compatibile con il budget già identificato dal DS e DG all'interno dei fondi di Astronomia Industriale (SKA-CTA) per investimenti su precursori SKA a bassa frequenza.**

Come secondo dato in TAB 2 riportiamo una stima di costi più articolata. In aggiunta ai costi di TAB. 1, questa seconda stima comprende anche il supporto al progetto in termini di infrastrutture di calcolo e analisi dati, di formazione per i giovani ricercatori, e di nuovo FTE da acquisire. **In particolare come discusso in Sez. 3.2 l'esistenza di una infrastruttura adeguata per l'analisi dati e un sostegno alla formazione dei giovani ricercatori sono CONDIZIONI STRETTAMENTE NECESSARIE alla buona riuscita dell'intera operazione.**

Nel caso della TAB.2 l'investimento è relativo all'intero consorzio LOFAR IT, al momento costituito da INAF e da UniTo (Sezione 4) ma potenzialmente espandibile.

In particolare i costi comprendono :

1. l'acquisto, installazione e gestione delle risorse di calcolo descritte in Sezione 3.2.2 e Appendice 1.

La TAB 2 **prevede una media Infrastruttura 1 affidata ad UniTo (in rosso)**. Questo sarebbe un investimento equivalente a circa 600 kEuro in 5 anni comprensivi di spese per energia elettrica e FTE. All'UniTo sarebbe richiesto di riservare/prioritizzare un certo numero di nodi di OCCAM del C3S per LOFAR IT (oppure di acquisirne nodi equivalenti o superiori) e una opportuna quota di archivio, cosa fattibile con fondi in gran parte già disponibili (Sez. 4).

La strategia suggerita è di **affidare ad INAF la gestione delle risorse di tipo 2**.

Al netto del personale, l'investimento stimato per 18 nodi di tipo 2 (Sez. 3.2.2, Appendice 1) è di circa 220 kEuro, tali risorse dovrebbero essere in grado di garantire una piena operatività fino a circa 20-25 utenti (staff, PhD, postdoc). Il modello suggerito in Sez.3.2.2 è di **6 nodi di tipo 2 installati presso 3 sedi INAF** con gruppi informatici ed esperienza adeguate alla gestione di tali risorse, in tal caso la gestione della risorsa non sarebbe tale da prevedere FTE aggiuntive (in blu in TAB 2). Inoltre nel caso di **risorsa distribuita i costi possono essere facilmente abbattuti sfruttando il co-finanziamento e/o auto-finanziamento di diversi gruppi di ricerca**, pur affidando ad INAF la gestione e operatività unica delle risorse. Ad esempio in TAB. 2 è evidenziato un possibile abbattimento dei costi di circa 110 kEuro (in blu), che tiene conto del fatto che 1 nodo di tipo 2 è già installato ad IRA (Appendice 1), che il gruppo di ricerca attivo nel campo degli ammassi di galassie dell'IRA prevede l'acquisto di altri 6 nodi (finanziati su fondi ERC e MAE) e infine che l'OA di Catania prevede l'acquisto di 2 nodi (finanziati su fondi esterni); tale finanziamento è sottratto in TAB. 2 per il calcolo delle spesa effettiva sostenuta da INAF (ultima riga).

2. E' stata prevista una FTE aggiuntiva da parte di INAF per il supporto informatico e l'ottimizzazione e sviluppo di pipelines. Tale FTE si combinerebbe con una FTE proveniente da UniTo e con FTE già disponibili all'interno di gruppi informatici di INAF per costituire un gruppo per il supporto informatico e lo sviluppo di pipelines di LOFAR IT (Sez.3.2.2, Appendice 1).
3. L'investimento tecnologico su LOFAR 2.0 già inserito in TAB 1 (Sez. 3.2.1, App 2).
4. Il costo della Stazione LOFAR 2.0 comprensivo di preparazione del terreno a Medicina, installazione ed operatività come da TAB 1 (Allegato 4).
5. I running cost annuali per la partecipazione al consorzio ILT (Allegato 3).
6. I costi per supportare il training di giovani ricercatori all'estero, stimati in circa 60 kEuro nei primi 2 anni del progetto (Sez. 3.2.3). Abbiamo anche previsto (in parentesi) i costi per un programma Nazionale di postdoc LOFAR come "desiderata". Costi per il finanziamento di borse di dottorato (Sez. 3.2.3) non sono conteggiati in TAB 2 essendo questi conteggiabili nell'ambito di accordi già finanziati e/o in itinere.

In TAB 2 abbiamo riassunto i costi valutabili in investimento TOTALE per anno (e sommati sui 5 anni) per l'intero consorzio e per INAF comprensivi anche degli FTE già disponibili in INAF. L'investimento INAF è valutabile in circa 3.6 MEuro in 5 anni (di cui 300 kEuro di postdoc) e quello di UniTo in 0.6 MEuro. Nelle ultime 2 righe della TAB 2 abbiamo indicato i costi (spese previste) per INAF basandoci su uno schema di desiderata, comprensivo di un programma Nazionale di postdoc (voci costo in parentesi in TAB 2) e su uno schema senza postdoc.

Nella stima dei costi abbiamo anche riportato (in \*) spese che possono essere valutate come contributo in-kind per ILT (es, sviluppo software) e che potrebbero essere detratte dai running costs previo accordo con l'ILT board.

TABELLA 2: PREVISIONE DI COSTI PER I PRIMI 5 ANNI

	2018	2019	2020	2021	2022	TOT
INFRASTR1	230					230
1FTE	45	60	60	60	60	285
RUNNING	15	20	20	20	20	95
RISORSA2	110	110				220
COFIN T2	-55	-55				-110
0.25FTExT2	20	45	45	45	45	200
SUPPORTO		60	60	60*	60*	240*
TECNO	20	90	40	10		160
FTE	30	90	90	30		240
STAZIONE				1650		1650
TERRENO				60		60
RUNNING					40	40
FTE				30	15	45
RUNNING	65	92	92	92*	92*	433*
TRAINING	25	35				60
POST DOC	(50)	(100)	(100)	(50)		(300)
TOT INV	610	702	507	2107	332	4258
INAF INV	320	622	427	2027	252	3648
INAF COST	270	487	292	1922	192	3163
				-60*	-60*	-120*
INAF MIN	165(215)	332(432)	192(292)	1872(1922)	192	2753(3053)
				-60*	-60*	-120*
						<b>2633/2933</b>

## 5.1 STRATEGIE PER LA DEI RIDUZIONE DEI COSTI.

Le TAB 1 e 2 delineano una roadmap fair dei costi da sostenere per l'ingresso di INAF in ILT e per l'ottimizzazione del ritorno scientifico e tecnologico dell'operazione.

Allo stesso tempo è ragionevole attuare delle politiche di ulteriore compressione della spesa per INAF, fra queste :

- i. coinvolgimento di ulteriori gruppi Universitari nel consorzio LOFAR IT. Questo attiverebbe un meccanismo di sharing dei costi (es running costs per ILT) e aprirebbe ad ulteriori canali di cofinanziamento sia delle risorse per l'analisi dati che del personale tecnologico (informatico) e scientifico (post-doc). Il coinvolgimento più ampio possibile della comunità Nazionale all'interno dei consorzi è nelle linee guida di ILT.
- ii. usando le attuali infrastrutture INAF (ad es HOTCAT, Appendice 1) per supportare parte dell'analisi dati LOFAR (es, parte della fase 2 di analisi che richiede meno RAM) riducendo o posticipando parte dell'investimento INAF nelle risorse di analisi dati;

- iii. Ottenendo un supporto dalla politica locale (es Comune di Medicina, Provincia etc) per i lavori di preparazione del terreno dove installare la Stazione.

## 5.2 CONCLUSIONI SUI COSTI

In conclusione LOFAR ha un costo per l'INAF nei prossimi 5 anni di circa 2.3-2.4 MEuro senza considerare le spese per le risorse di calcolo e analisi dati, per training e per nuovo personale. Queste voci aggiuntive richiedono un finanziamento fino a circa 500 kEuro su 5 anni al netto dei co-finanziamenti attualmente in essere. Riteniamo NECESSARI almeno gli investimenti sulle risorse di calcolo e training di giovani ricercatori.

La spesa complessiva è sostanzialmente ripartita in **infrastrutture** (circa 1.85 MEuro fra Stazione LOFAR e calcolo), **running cost** per la partecipazione al consorzio per 5 anni (circa 0.4 MEuro), spese di supporto alla ricerca tecnologica (circa 0.2 MEuro) e spese per **nuovo personale** informatico/scientifico (circa 0.4 MEuro).

Questo computo è ottenuto considerando una tipologia minima di consorzio LOFAR IT basato su INAF e UniTo. Una significativa riduzione dei costi, principalmente per la voce running cost e nuovo personale deriverebbe dall'allargamento del consorzio ad altri gruppi Universitari.

# APPENDICE 1 : INFRASTRUTTURA ANALISI DATI E RETI

A. Bonafede, U. Becciani, M. Nanni, G. Taffoni, (G. Brunetti)

## A1.1. INTRODUZIONE

In questo studio sono state analizzate le richieste tecnologiche per l'analisi dati LOFAR (pipeline software che viene utilizzata per il data analysis, richieste computazionali e storage requirements) allo scopo di proporre una infrastruttura modulare e con crescita progressiva e una roadmap ad essa collegata, che permetta la realizzazioni dell'infrastruttura in tempi medio/brevi. Lo studio è fatto tenendo anche in considerazione le possibilità finanziarie per la costituzione completa dell'infrastruttura, e di un'utenza di 20-25 ricercatori che avranno necessità di analisi dati del progetto stesso.

L'infrastruttura di calcolo proposta da INAF fa inoltre parte dell'infrastruttura di calcolo globale del Consorzio Italiano per la partecipazione Italiana a LOFAR e quindi il suo disegno tiene anche conto di possibili investimenti pianificati da altri membri del consorzio (ad oggi Università di Torino).

La Stazione LOFAR 2.0 sarà presumibilmente installata entro il 2021, ma l'ingresso di INAF in ILT è previsto da subito e l'INAF con la sua infrastruttura di calcolo, gestirà l'analisi dati e l'archivio dei dati generati dai suoi gruppi.

In una fase successiva si può anche ipotizzare che LOFAR IT si faccia carico anche di un centro di storage per l'archivio LOFAR ILT, attualmente formato da 3 centri in Europa per complessivi 30 PByte circa (FIG 3). In tale fase sarebbe auspicabile poter esporre come INAF alcuni PByte (circa 10).

Un piano di finanziamento per la creazione di una infrastruttura di calcolo, limitandoci ai primi 2-5 anni, è proposto alla fine di questa Appendice.

## A1.2. STATO DELL'ARTE DEL COMPUTING IN INAF

Come primo passo è stata fatta una ricognizione degli investimenti recenti da parte di INAF nell'area del calcolo sia in regime HPC che in HTC. Inoltre molte sono le iniziative di sedi locali che hanno incrementato le risorse computazionali dipartimentali finalizzate esclusivamente a progetti scientifici di rilevanza nazionale o internazionale come per esempio: Alma Regional Center, Euclid presso Osservatorio di Trieste, Gaia e il DPCT di Altec di Torino e altre esperienze legate ad altri progetti. Questa ricognizione è utile per capire se esistono attualmente in INAF risorse computazionali per l'analisi di dati LOFAR. Infine riguardo a LOFAR segnaliamo che in INAF (IRA) esiste un primo nodo di calcolo acquisito nel 2017 su fondi Europei.

Allo stato attuale si possono distinguere due principali iniziative finanziate direttamente dalla Direzione Scientifica tramite coordinamento dell'ICT in ambito HPC e HTC: il Mou INAF-Cineca e il progetto CHIPP.

### A1.2.1 Il Mou del CINECA

L'utilizzo di una facility di tipo HPC Tier-0 (capacità di calcolo dell'ordine dei 10 PetaFLOPs) è necessaria per rispondere alle sfide che i grandi progetti di terra e di spazio esigono. In particolare SKA e CTA, ma anche Euclid, LOFAR, GAIA sono esempi di progetti in cui HP data analysis e data reduction (HPDA) si rendono necessari per fornire risultati scientifici alla comunità. Quindi anche i ricercatori INAF che usano precursori o simulazioni, al fine di mantenere i livelli di eccellenza raggiunti nelle loro ricerche, hanno necessità di accedere a risorse di tipo Tier-0.

L'INAF ha sottoscritto un MoU con il Cineca per il triennio 2017-2019 per l'uso dell'infrastruttura per un impegno complessivo di circa 360 kEuro.

### **A1.2.2 Infrastruttura di calcolo offerta dal CINECA**

INAF ha accesso a fino 50 Milioni di ore di calcolo per anno su sistemi Tier-0 e Tier-1 del Cineca. L'infrastruttura principale è il sistema Marconi con 3 partizioni:

1. System A1 (Broadwell), Modello: Lenovo NeXtScale, Nodi: 1.512, Processori: 2 x 18-cores Intel Xeon E5-2697 v4 (Broadwell) a 2.30 GHz, Cores: 36 cores/node, 54.432 core in totale RAM: 128 GB/node, 3.5 GB/core Peak Performance: 2 PFlop/s
2. System A2 (Knights Landing), Modello: Lenovo Adam Pass, Nodi: 3.600 Processori: 1 x 68-cores Intel Xeon Phi 7250 CPU (Knights Landing) a 1.40 GHz Cores: 68 cores/node (272 con HyperThreading), 244.800 core in totale, RAM: 16 GB/node of MCDRAM and 96 GB/node DDR4, Peak Performance: 11 PFlop/s.
3. System A3 (Skylake), Modello: Lenovo Stark, Nodi: 1.512 + 792, Processori: 2 x 24-cores Intel Xeon 8160 CPU (Skylake) a 2.10 GHz, Cores: 48 cores/node 72.576 + 38.016 cores in totale, RAM: 192 GB/node DDR4, Peak Performance: 7.00 PFlop/s

### **A1.2.3 Il progetto pilota CHIPP**

La Direzione Scientifica si è posta l'esigenza di soddisfare le necessità di calcolo di una fascia di utenza che ha bisogno di una infrastruttura di tipo Tier-2. Esigenza emersa in tutti i meeting di tutte le macroaree che si sono svolti sia nel 2016 che nel 2017.

Si tratta di rispondere alle esigenze di progetti di ricerca che fanno uso in prevalenza di codici debolmente paralleli o imbarazzantemente paralleli, con richieste per il singolo nodo di molta memoria RAM (superiore ai 128 GB); anticipiamo che questa è una configurazione di interesse in ambito LOFAR. Inoltre CHIPP si propone per supportare ricercatori che hanno necessità di fare sviluppo, debugging o profilazione di codici paralleli ad alte prestazioni, o di progetti scientifici che nelle fasi iniziali della loro vita hanno bisogno di fare test o analisi su infrastrutture di calcolo HPC o HTC.

Il progetto Pilota CHIPP di durata biennale, finanziato dalla Direzione Scientifica, ha lo scopo di soddisfare tali esigenze che difficilmente hanno spazio nell'infrastruttura di tipo Tier-0 e Tier-1 del Cineca.

Sulla base delle esperienze maturate nell'offerta di servizi di calcolo e delle disponibilità di infrastrutture già esistenti in INAF (acquisite tramite altri progetti), la Direzione Scientifica ha finanziato il progetto CHIPP valorizzando i centri di calcolo di Catania e di Trieste. Altri centri potranno eventualmente aggiungersi agli esistenti.

L'investimento fatto ammonta a circa 200 kEuro .

Sono garantiti ai ricercatori INAF fino a 3.5 milioni di ore per anno di calcolo su :

1. HOTCAT (INAF-OATS)  
Nodi 20 con 40 Core INTEL Haswell E5-4627v3 @ 2.60GHz (4 SOCKET); 6GB RAM/Core (256 GB RAM per nodo). Risorsa disponibile per CHIPP pari al 40% del tempo totale.
2. CLOUDCAT (INAF-OATs)  
è un sistema cloud basato su software OpenStack (Mikata). Nodi 12 con INTEL Westmere E5620 @ 2.40GHz 16 core con tecnologia HyperThreading (8 core e 8 thread); 8GB RAM/Core (64GB RAM totale per nodo). Risorsa disponibile per CHIPP pari al 70% del tempo totale.
3. MUP (INAF-OACT)  
Nodi 16 con 12 Core (6 Hyper-Threading) Intel® Xeon® E5-2620@3.2; 5.2GB RAM/Core (64GB RAM totale per nodo). Risorsa disponibile per CHIPP pari al 90% del tempo totale.

#### **A1.2.4 Elaborazione dati radioastronomici all'IRA ed Alma Regional Center**

Da 10 anni l'IRA ospita l'Alma Regional Center ed ha cercato di trovare le soluzioni hardware piu' adeguate per gestire i dati e l'applicazione CASA, software sviluppato appositamente dal consorzio NRAO - ESA - NAOJ per gestire i dati del radiotelescopio Alma.

Le scelte hardware fatte dall'ARC nel 2008 si basavano su specifiche tecniche della applicazione CASA che prevedevano un medio (64-128) grado di parallelismo delle applicazioni maggiormente time consuming. Per questo si e' realizzato un cluster basato su blade server DELL capace di ospitare 16 lame (96 core in totale) . Attualmente sono attive le seguenti risorse:

1. Arcbl01-12 (INAF-IRA)  
Nodi 12 con 8 Core AMD Opteron(tm) Processor 2352@2.1GHz 1GB RAM/Core (8GB RAM totale per nodo).
2. Arcbl13-14 (INAF-IRA)  
Nodi 2 con 16 Core AMD Opteron(tm) Processor 2387@2.8GHz 2GB RAM/Core (32GB RAM totale per nodo).

Il blade server DELL e' collegato in ethernet a 10Gbit/s con un filesystem LUSTRE basato su 2 DTN hosts per una capacita' complessiva di circa 100 TByte.

Questa configurazione e' attualmente in discussione in quanto, nonostante le aspettative, il parallelismo nel software CASA e' stato sviluppato solo per alcuni dei moduli e non ha portato vantaggi significativi, e soprattutto si e' scoperto che, nonostante il filesystem LUSTRE sia particolarmente efficiente, a causa dell'uso condiviso risulta lento in confronto al caso di dischi interni ai singoli server. Per contro i dischi interni, cioe' collegati direttamente sul bus dei sistemi, risultano insufficienti per capacita (<1TBy) in molte delle applicazioni.

Si sta' quindi migrando su una architettura a FAT-STATION, cioe' sistemi indipendenti, dotati di 4-8 dischi in raid, che condividono lo spazio /home e lo storage LUSTRE, ma che prevedono che i singoli progetti siano elaborati sui dischi locali. Questa architettura, gia' utilizzata per la riduzione di dati con AIPS, e' attualmente in fase di test per applicazioni ARC e LOFAR sui seguenti sistemi.

- 1) LOWCOST: FAT-STATION (INAF-IRA)  
Nodo con 4 Core AMD Athlon(tm) X4 860K @3.4GHz 8 GB RAM/Core (32 GB RAM totale). Disco locale 11TBy. Risorsa per test delle applicazioni ARC
- 2) LOFAR: FAT-STATION (INAF-IRA)  
Nodo con 40 Core (20 Hyper-Threading) Intel® Xeon® E5-2640@3.4GHz (2 SOCKET), 512 GB RAM totale. Disco locale 28TBy.  
Risorsa per test e sviluppo applicazioni LOFAR

Si sta' anche considerando di utilizzare di una tape library per non mantenere on-line i dati scarsamente utilizzati.

#### **A1.3. REQUISITI PRELIMINARI SOFTWARE E HARDWARE PER ANALISI DATI LOFAR**

Lo studio ha preso come riferimento un tipico dataset prodotto dalla survey Tier-1 (Sez. 1.2 e 3.2.1) e le ultime versioni delle pipelines per l'analisi delle osservazioni LOFAR HBA nel continuo prodotte dal SKP e da ASTRON.

Per determinare i requisiti preliminari software e hardware dell'infrastruttura, si individuano le principali fasi nell'analisi dei dati.

**A1.3.1 Pre-factor (fase 1.a)** (flag - calibrazione del calibratore - trasferimento delle soluzioni al target e calibrazione iniziale del target - averaging dati).

Dopo aver rimosso dati spuri e mediato in frequenza e in tempo, i dati del calibratore vengono calibrati tramite una pipeline automatica. Tipicamente l'utente esegue un check della calibrazione durante le fasi di esecuzione. Il check da parte dell'utente richiede un accesso interattivo alla risorsa di calcolo da effettuare con tecnologie del tipo remote desktop. La calibrazione del calibratore deve essere fatta prendendo tutti i dati, ovvero senza divisione in intervalli di frequenza (bande). Vengono prodotti automaticamente dei plot per verificare la qualità della calibrazione. Successivamente, alcuni risultati della calibrazione vengono trasferiti al target (gain di ampiezza e clock-offset), e il target viene calibrato utilizzando un modello a bassa risoluzione estratto dalla survey TGSS. Poiché l'osservazione del target è molto più lunga di quella del calibratore (circa 8 ore contro i 15-30 min del calibratore) è possibile dividere le osservazioni del target in bande di frequenza e calibrare indipendentemente queste bande (di solito vengono calibrati insieme 10-12 sottobande per un totale di 2 MHz di banda). Un'osservazione tipica produce quindi alla fine di questa fase 25 bande ognuna di 2 MHz di banda che sono l'input della fase di initial-subtraction (fase 1.b). Pre-factor (fase 1.a) richiede circa 200 GB di RAM ed è limitata da I/O. I dati in input sono 10-20 TB (suddivisi in files da 64GB l'uno) e in output circa 500 GB. Tempo calcolo medio stimato prendendo come riferimento le macchine LOFAR operative a Leiden (n. 2 Intel Xeon E5-2697A v4 16 core a 3.2GHz, 32 threads, con 256 GB Ram con capacità di 563 GFLOPS): ~ 3-4 giorni. L'aspetto limitante per le prestazioni è la quantità di RAM per nodo (a livello ottimale per core), dato che i dati sono splittati in file da 64 GB la quantità di memoria richiesta per analizzare il singolo file sarà dell'ordine di 64GB.

**A1.3.2 Initial-subtraction (fase 1.b)** (immagini iniziali a bassa e alta risoluzione, dataset "vuoto" e modelli di auto-calibrazione che sono l'input di factor - fase 2)

Questa fase serve per preparare i dataset alla fase factor (fase 2) e allo stesso tempo vengono prodotte immagini dell'osservazione che hanno una prima calibrazione (effettuata in pre-factor). Viene fatta un'immagine dell'osservazione ad alta risoluzione che serve per identificare le sorgenti da sottrarre ai dati, e che andranno a formare il modello di auto-calibrazione ad alta risoluzione per factor. Una volta che le sorgenti identificate ad alta risoluzione vengono sottratte dai dati, vengono fatte nuove immagini a risoluzione più bassa per identificare le sorgenti più estese e a distanza maggiore dal centro del puntamento dell'osservazione. Le immagini in questa fase sono tipicamente di circa 6000 x 6000 pixels.

Initial-subtraction (fase 1.b) è RAM intensive, con picchi di 400GB. I dati in input sono circa 500 GB e l'output produce circa 80 GB in più. Tempo calcolo stimato prendendo come riferimento il nodo LOFAR presso l'IRA (Sez. A1.2.4): 4-5 giorni

**A1.3.3 Factor (fase 2)**

In questa fase la calibrazione procede per diverse regioni di cielo, in modo da correggere effetti locali di cui la calibrazione iniziale non tiene conto. Il campo di vista viene diviso in diverse facets a seconda degli input dell'utente, e si procede alla calibrazione delle facet in maniera indipendente. Si possono calibrare insieme su più nodi di calcolo in maniera "parallela" per velocizzare il processo. La calibrazione e le immagini che seguono vengono ripetute più volte cambiando i parametri a seconda della regione di cielo da analizzare e degli obiettivi scientifici dell'osservazione. **Factor (fase 2) è computing intensive**, prende in input i dati e i modelli creati dalla fase di initial subtraction. Per non introdurre danni irreparabili ai dati, questi vengono copiati prima di passare alla calibrazione. RAM richiesta: dipende dalla dimensione delle facets. Tipicamente la RAM richiesta è <128 GB ma si può arrivare fino a 256 GB in casi particolari. L'output è circa 200 GB, ma dipende molto dal tipo di osservazione. Tempi: mese.

Dallo studio del software di analisi dati di LOFAR si evince che esistono alcuni requisiti hardware principali: livello di RAM richiesto, tipologia filesystem, connessione di rete tra i nodi e i centri di calcolo, tipologia di CPU.

#### **A1.4. CONNESSIONE DI RETE GEOGRAFICA DEI NODI DI CALCOLO**

La fase iniziale di analisi, come precedentemente descritto, richiede l'iniziale trasferimento di dati dagli archivi di LOFAR di una quantità di dati che è dell'ordine di 10-20 TB per osservazione di tipo Tier-1, fino a 100 TB per osservazioni di tipo Tier-2 o 3 (Sez. 1.2). La quantità di dati apparentemente non è proibitiva ed è ragionevole attendersi che, data una connessione non dedicata a 1 Gbit/s siano necessarie circa 4 ore per trasmettere 1 TBy di dati.

Esperienze sul campo dimostrano che il trasferimento di tipo "general internet" tra nodi astronomici a livello europeo riescono a fornire velocità di una o due ordini di grandezza inferiori. E' quindi necessario che i nodi INAF che partecipano alla rete di calcolo siano dotati di connessioni di tipo Sci.DMZ (cioè che non siano limitati da politiche di firewall) e che almeno uno dei nodi di calcolo INAF sia visto come tale dal LOFAR data storage e che non venga considerato come un nodo "general internet".

E' auspicabile che i nodi di calcolo che parteciperanno alla rete di calcolo siano collegati con banda da 10 Gbit/s. In termini di costo questo significa una spesa di circa 10 KEuro per gli apparati (switch-router) e un incremento di circa 17 KEuro/anno per i collegamenti in fibra che attualmente sono a 1 Gbit/s. ( da 10 KE/anno a 27 KEuro iva inclusa). Essendo di interesse comune per una serie di progetti futuri in cui INAF è già coinvolta, questo costo non sarà conteggiato per i costi relativi all'investimento in LOFAR (Sez. 5).

#### **A1.5. RICHIESTA NODI DI CALCOLO**

##### **A1.5.1 Pre-factor (Fase 1.a)**

Il processo di pre-factor dal punto di vista computazionale e' un processo imbarazzantemente parallelo potendo operare su diversi data files da 64GB contemporaneamente. Tale processo richiede una ram di 64 GB per core e almeno 8 core per nodo, alternativamente un nodo costituito da un numero maggiore di cores ma con una RAM complessiva di 512-1000 GB; tale seconda opzione è più costosa e porta ad un uso meno efficiente dei cores in un nodo. Il nodo di calcolo in questa fase e' dedicato esclusivamente al run di pre-factor.

In questa fase la pipeline prevede un'unica interazione da parte dell'utente che deve controllare la calibrazione del calibratore prima di applicarla al target possibilmente attraverso sistemi di remote desktop. La procedura è scritta in linguaggio Python (prevalentemente) e organizzata in step sequenziali in particolare un calibrator step e un target step.

##### **A1.5.2 Initial-subtraction (Fase 1.b)**

Questa fase è finalizzata alla generazione di: 1) immagini iniziali di tutto il campo di vista dello strumento, 2) un modello delle sorgenti che serve come input a factor, 3) un dataset virtualmente "vuoto" che serve come input a factor (fase 2). Questa fase e' tipicamente eseguita una sola volta per data set.

I dati di ingresso che vengono utilizzati sono i risultati della fase di pre-factor che corrispondono a ~500GB. Si tratta di una pipeline seriale che concatena programmi scritti in linguaggi diversi (principalmente python, C e C++). I task possono essere eseguiti in parallelo con tecnologia multi-threading e coinvolgono un singolo nodo. Per questa fase e' necessaria una memoria di circa 70-80 GB RAM per banda da elaborare, con un picco di circa 400 GB per un periodo di circa 4 - 5 ore del tempo totale di elaborazione (dai test effettuati sul nodo LOFAR ad IRA). Il tempo di calcolo della fase di initial-subtraction e' di circa 4-5 giorni su

nodo con capacità di circa 500 GFLOPS . Dato che per questa fase si richiede un nodo con almeno 512 GB di RAM, il nodo non può essere condiviso con altri tasks ma completamente dedicato; eventuali condivisioni richiedono nodi con 1+ TB di RAM.

### A1.5.3 Factor (Fase 2)

E' la fase che tende a impegnare la maggior quantità di tempo per l'utente. Può essere ripetuta più volte, e con check fatti durante il processamento al fine di verificare lo stato di esecuzione e controllare i dati preliminari; questa fase richiede una notevole interattività da parte dell'utente che non solo verifica il processo ma può anche decidere di interrompere e far ripartire l'analisi più volte.

Si tratta di una pipeline parallelizzabile con paradigma "embarrassing parallel" con step principalmente scritti in Python e C++ che parte dai ~500GB di dati prodotti dalla fase 1.b. Il livello di parallelismo su più nodi si raggiunge dividendo il cielo osservato in più regioni. All'interno di ciascuna regione del cielo è possibile parallelizzare ulteriormente l'analisi eseguendo il software in modalità multithreading operando su singolo tile (n.b. non più di 10 thread per nodo). Questo tipo di analisi e' **computing intensive e coinvolge da 16 a 24 cores distribuiti** anche su più nodi, ha invece un uso intenso della RAM rispetto alla Fase 1, dato che l'intera pipeline richiede meno di 256 GB di memoria. Il tempo di esecuzione medio può variare da un minimo di una settimana fino a diverse settimane a seconda delle interazioni e delle scelte dell'utente.

### A1.5.4 Considerazioni generali sul filesystem

In tutte e tre le fasi sopra descritte, l'accesso ai dati rappresenta, assieme alla RAM, il fattore limitante principale nell'esecuzione dei tasks , quindi molta attenzione deve essere dedicata al setup dello storage e alla sua ottimizzazione. Inoltre, per come sono costruite le pipeline, e' necessario che i nodi che partecipano alla riduzione abbiano un filesystem condiviso ad alte prestazioni oltre che a dei dischi locali veloci eventualmente di tipo SSD o SAS. La modalità di condivisione (e.g. NFS, parallel filesystem o Object storage) e' particolarmente importante per ottimizzare l'I/O.

Una soluzione e' rappresentata dall'uso di un filesystem parallelo di cluster che garantisca sia maggiori prestazioni che scalabile con il tempo. Ad esempio BeeGFS permette di implementare una alta ridondanza dei dati (se necessario) con delle repliche automatiche ad alta granularità (e.g. ridonando un singolo file o una singola directory). Inoltre e' scalabile per cui si possono aggiungere nodi di storage (o anche dischi) mantenendo attivo il filesystem. In un'ottica di ottimizzazione, gli stessi nodi di calcolo possono diventare dei nodi di storage partecipando con i loro dischi locali alla creazione del filesystem parallelo, questo a patto di riservare almeno due cores e 50 GB di RAM per la gestione del filesystem. Un filesystem parallelo tipo Beegfs e' in grado di raggiungere i 100Gb/s in lettura.

Una possibile alternativa dalle prestazioni simili ma in grado di scalare in modo più efficiente sono i cosiddetti Object storage, in particolare quelli che garantiscono una interfaccia con semantica POSIX. Un esempio e' CEPH filesystem che e' in grado di erogare un sustained throughput in lettura di circa 100 Gb/s con un sistema composto da 24 HDD SATA. Un sistema object storage e' scalabile e facilmente mantenibile e garantisce delle interfacce per l'upload e download dei files dall'esterno basate anche su tecnologie web (e.g. REST APIs).

Qualunque sia la scelta implementativa, il filesystem dovrà essere dimensionato in modo da permettere ai gruppi di lavoro previsti dal progetto di lavorare contemporaneamente, quindi deve prevedere uno storage complessivo Tier 1+2 di almeno 500TB.

### A1.5.5 Considerazioni generali

Dato che una significativa parte della data reduction e analisi richiede interattività da parte dell'astronomo, è necessario facilitare l'accesso remoto ad infrastrutture di calcolo. Una

possibile soluzione è implementare nei diversi centri di calcolo un sistema di remote desktop che garantisca una efficace compressione dei dati (in particolare grazie ad hardware dedicato) e che quindi possa permettere un'esperienza fluida di accesso alle risorse computazionali; questo è particolarmente utile per la Fase 2.

Dato che i sistemi di calcolo necessari per supportare le operazioni di riduzione e analisi dei dati LOFAR sono condivise tra diversi gruppi e al fine di ottimizzarne l'uso, e' necessario che tali risorse siano configurate come un sistema di calcolo multi-utente e quindi implementino un resource manager (queue system e scheduler).

Nel caso di una organizzazione delle risorse secondo un modello distribuito (come spiegato in dettaglio nel seguito del documento), diventa molto importante che le risorse stesse abbiano un ambiente software identico e che venga implementato un sistema efficiente per il deployment e mantenimento del software scientifico (LOFAR pipelines). Le tecnologie più promettenti allo stato attuale sono basate sull'uso di "containers". Un container e' un sistema per automatizzare il deployment di applicazioni all'interno di contenitore software, fornendo un'astrazione aggiuntiva grazie alla virtualizzazione a livello di sistema operativo di Linux. Essi utilizzano le funzionalità di isolamento delle risorse del kernel Linux, come ad esempio cgroups e namespaces, per consentire a "container" indipendenti di coesistere sulla stessa istanza di Linux, evitando l'installazione e la manutenzione di una macchina virtuale. In ambiente di calcolo distribuito o HPC e' largamente utilizzato un sistema di container denominato "Singularity " che permette tra le altre cose di condividere delle immagini contenenti il software di data reduction e l'ambiente linux in modo che le varie risorse di calcolo appaiono perfettamente omogenee agli utenti.

Tale tecnologia insieme ai container Docker è in fase di studio da IRA, OATs e OACt per effettuare test su infrastrutture diverse (Bologna IRA, CHIPP Trieste e Catania).

Al fine di facilitare l'accesso alle risorse di calcolo è necessario predisporre un sistema di autenticazione federata, ad esempio attraverso un sistema di tipo LDAP gestito unitariamente dal consorzio e sotto la responsabilità tecnica di uno staff tecnico del consorzio di cui si auspica la formazione.

Come anticipato in questa relazione ci occupiamo di definire una strategia tecnica per l'ottimizzazione infrastrutturale dei primi 2-5 anni dall'ingresso di INAF in ILT. In tale fase non è richiesto un grande storage permanente (e quindi incrementale) dei dati utilizzati dato che il sistema non farebbe parte dell'archivio ILT. E' tuttavia auspicabile che il sistema computazionale a cui partecipa INAF abbia la possibilità di archiviare i prodotti avanzati ottenuti dalla fase di Factor in modo permanente come archivio locale o storage che sia disponibile ai vari gruppi di ricerca. In tal caso sarà importante il coinvolgimento del gruppo del centro italiano archivi IA2 e ICT.

#### **A1.6. INFRASTRUTTURA ITALIANA**

Il gruppo di lavoro raccomanda l'adozione del modello MONARC anche se non rigidamente inteso, cioè con possibilità dinamica di sovrapposizione di compiti e task specifici dei vari livelli di Tiers . E' necessario che l'intera infrastruttura sia in grado di soddisfare un'utenza media prevista di 20-25 utenti, geograficamente distribuiti nel territorio Italiano.

Si propone il sistema distribuito su 2 livelli di tiers:

- ❖ Tier-1, una infrastruttura unica, ubicata presso un membro del consorzio LOFAR IT e in grado di soddisfare l'utenza per le fase 1a e 1b, ed eventualmente contribuire in modo significativo all'archiviazione dei prodotti avanzati dell'analisi dati;
- ❖ Tier-2, una infrastruttura dedicata alla fase 2 dell'analisi dati. Sulla base del nostro studio il processo di analisi dati non dipende in modo sensibile dal fatto che questa infrastruttura sia unica oppure distribuita sul territorio. Riteniamo tuttavia che un

Tier-2 distribuito su scala nazionale (es 3 poli) minimizzi i costi per INAF (hardware e FTE) sul breve-medio termine.

#### **A1.6.1 TIER-1:** destinato prevalentemente alla realizzazione delle fasi 1.a e 1.b

##### Requisiti Hardware consigliati

- Cluster di Calcolo costituito da nodi FAT con disponibilità di RAM di 512 GB fino ad 1TB. 1 o 2 Socket con processore Intel Xeon di ultima generazione o equivalenti e, ad esempio un numero di core pari ad 8 o 16 con 64 GB RAM/core, alternativamente un numero maggiore di core/nodo con richiesta proporzionalmente inferiore di RAM/core. Interconnessione su rete infiniband 56 Gbs (o superiore).
- Storage distribuito ad alte prestazioni di almeno 80 TB eventualmente basato su dischi SSD o SAS.
- Un secondo sistema di storage distribuito a basse prestazioni fino a circa 500 TB-1 PB.
- Remote desktop per le necessità descritte in precedenza.
- Connessione alla rete GARR da almeno 10 Gbps.

##### Software suggerito:

- Sistema Operativo, Linux
- Schedulers and Workload Management, a questo scopo possono essere considerati validi candidati PBSPro, Torque, LSF ecc ed altri sistemi con analoghe caratteristiche.

Devono essere presi in considerazione gli aspetti di gestione legati alla scalability e alla latenza:

- pianificazione basata su policy di bilanciamento dei tempi e l'utilizzo delle risorse con una distribuzione ottimale;
- resilienza con sistemi fail-over automatica in modo che i jobs della pipeline non vengono mai persi anche in caso di failures;
- Flexible Plugin Framework per permettere estensibilità, personalizzazioni e soddisfacimento di requisiti complessi;
- health checks, con monitoraggio automatico e con un quadro completo di controllo.
- sistemi di monitoraggio del carico e del suo bilanciamento.
- Distribuzione del software scientifico. Nell'ipotesi che le procedure della fase 1 possano essere gestite in modo efficiente con l'uso dei container, si configura come ideale la possibilità di utilizzare singularity o docker

In conclusione, per soddisfare l'utenza contemporanea di almeno 3-5 utenti scientifici, si deve prevedere la disponibilità di almeno 3 nodi dedicati a tale fase.

#### **A1.6.2 TIER-2:** destinato prevalentemente alla realizzazione della fase 2 (factor)

##### Requisiti Hardware consigliati

- Cluster di Calcolo costituito da nodi con disponibilità di RAM di almeno 256 GB. 2 Socket con processore Intel Xeon di ultima generazione o equivalenti e un numero di core, ad esempio pari ad almeno 8-16 per socket con 16 GB RAM/core. Interconnessione su rete infiniband 56 Gbs (o superiore); alternativamente un numero maggiore di core/nodo con minore richiesta di RAM/core.
- Storage distribuito ad altissime prestazioni di almeno 10 TB eventualmente basato su dischi SSD o SAS.
- Un secondo sistema di storage distribuito a basse prestazioni di almeno 50 TB.
- Remote desktop per le necessità descritte in precedenza.
- Connessione alla rete GARR ad almeno 1 Gbps (auspicabili a 10 Gbps).

##### Software suggerito

- Sistema Operativo, Linux

- Schedulers and Workload Management, a questo scopo possono essere considerati validi candidati PBSPro, Torque, LSF ecc ed altri sistemi con analoghe caratteristiche. Devono essere presi in considerazione gli aspetti di gestione legati alla scalability e alla latenza:

- pianificazione basata su policy di bilanciamento dei tempi e l'utilizzo delle risorse con una distribuzione ottimale;
- resilienza con sistemi fail-over automatica in modo che i jobs della pipeline non vengono mai persi anche in caso di failures;
- Flexible Plugin Framework per permettere estensibilità, personalizzazioni e soddisfacimento di requisiti complessi;
- health checks, con monitoraggio automatico e con un quadro completo di controllo.
- sistemi di monitoraggio del carico e del suo bilanciamento.
- Distribuzione del software scientifico. Nell'ipotesi che le procedure della fase 2 possano essere gestite in modo efficiente con l'uso dei container, si configura come ideale la possibilità di utilizzare singularity o docker

Nota: il dimensionamento di alcuni nodi del sito Tier-2 con memoria di 512 GB potrebbe anche permettere l'esecuzione della fase 1 sul sito. Questo lo si ritiene auspicabile anche come sistema di riserva in caso di indisponibilità temporanea del Tier-1 anche in caso di occupazione programmata per lavori di altri gruppi di ricerca. La sovrapposizione dei compiti da Tier-1 a Tier-2 non è una soluzione da adottare in fase operativa come regola di operazione, ma è da considerare in casi di effettiva necessità.

Per soddisfare l'utenza di diversi gruppi di ricerca (fino a circa 20-25 utenti attivi), considerato che la fase 2 dura per diverse settimane, è consigliato avere almeno 3 siti distinti dedicati ciascuno dotato di 6 nodi di calcolo e relativi storage. Per rispondere anche alla distribuzione geografica dei gruppi di ricerca è auspicabile che i 3 siti coprano le 3 aree geografiche del paese, Nord, Centro e Sud e soprattutto che i siti trovino collocazione dove maggiore è stata la competenza e la gestione di risorse di calcolo per la comunità.

### **A1.6.3 infrastruttura generale**

Al fine di garantire un accesso trasparente agli utenti nei centri afferenti all'infrastruttura, e' necessario implementare un sistema di autenticazione federato o distribuito (e.g. Ldap o similari) che tenga conto che e' necessario un accesso di tipo shell alle macchine di calcolo.

Attività minima di coordinamento per garantire un monitor e controllo della fase operativa di tutti i siti appartenenti alla infrastruttura complessiva (Tier-1 e Tier-2).

Un punto cruciale e' la gestione del software scientifico che deve essere omogeneo tra tutti i siti e aggiornato con continuità. Per questo si devono prevedere sistemi automatici per la installazione usando container o tecnologie simili ad esempio il CVMFS .

## **A1.7. STRUTTURA ORGANIZZATIVA**

Si ritiene necessaria anche la formazione di un gruppo tecnico che possa dettagliare i requisiti dell'infrastruttura, mantenere il know how tecnologico aggiornato, mantenere e gestire l'infrastruttura, gestire l'accesso federato, monitoraggio dell'uso delle risorse con evidenziazione dei punti di criticità e strategia per il miglioramento del servizio offerto ove necessario. Un investimento in termini di personale qualificato da inserire in tale gruppo aprirebbe anche le porte alla possibilità di sviluppare e/o ottimizzare pipelines per la riduzione dei dati LOFAR dando un peso maggiore a LOFAR IT nell'ambito di ILT. Tale investimento potrebbe anche essere conteggiato come contributo in-kind alla partecipazione di ILT (Sez.5).

## **A1.8. INVESTIMENTO PROPOSTO PER INAF IN INFRASTRUTTURA**

La grande richiesta di RAM/nodo necessaria all'analisi dati LOFAR richiede una infrastruttura di calcolo molto specifica. **Sulla base del nostro studio concludiamo che ad oggi in INAF non esistono risorse appropriate per l'analisi dati LOFAR fatto salvo un unico FAT node di calcolo installato presso l'IRA (Sez. A1.2.4).** L'unica altro cluster che al momento potrebbe essere utilizzato per l'analisi dati LOFAR, almeno per la fase 2 o parte di essa, è HOTCAT a Trieste (Sez. A1.2.3) che tuttavia non è un cluster dedicato ed è condiviso da numerosi utenti provenienti da diversi progetti.

Dalla studio fatto e sopra riportato, e in attesa anche di maggiori specifici risultati di test che sono in fase di esecuzione a Bologna, Trieste e Catania, si ipotizza una infrastruttura articolata del consorzio e di INAF in particolare, ripartite nelle tre macro-regioni italiane (Nord, Centro e Sud).

**Il sito primario, hosting del Tier-1** basato su almeno 4 nodi di calcolo FAT (fino a 1 TB di RAM/nodo) e con uno storage di diversi x100 TB, avrà un costo di investimento iniziale di almeno 300 kE. A questo costo bisogna aggiungere le spese di management 1,5 FTE di consumo elettrico, dell'ordine di 20 kEuro circa.

**Un tipico sito di Tier-2** composto da 6 nodi di calcolo, con infrastruttura di rete e di storage, così come delineato nel par 6.2 ha un costo di acquisto di circa 80- 100 kEuro.

A questo costo bisogna aggiungere le spese di management 0.5 FTE almeno per sito e di consumo elettrico, dell'ordine di 8 kEuro circa.

Per ottimizzare i costi e facilitare lo startup, è opportuno valorizzare le infrastrutture di calcolo già disponibili e che abbiano già maturate competenze nella gestione dei sistemi di calcolo.

## APPENDICE 2: COLLABORAZIONE TECNOLOGICA SU LOFAR 2.0 & STAZIONE A MEDICINA

P. Bolli, J. Monari, M.Nanni, F.Perini, (G. Brunetti)

La comunità tecnologica INAF è attiva nel campo degli Aperture Arrays a bassa frequenza (LFAA) in ambito SKA sin dal 2004, prima col progetto SKADS e poi dal 2010 al 2013 con l'Aperture Arrays Verification Program (AAVP). Oggi è coinvolta nell'Aperture Arrays Design Consortium (AADC), consorzio di cui coordina il Work Package sui ricevitori (WP-RX), e che dal 2014 si occupa della progettazione dell'array a bassa frequenza (LFAA) per SKA.

Al gruppo tecnologico LFAA di INAF afferisce personale da diversi Istituti con diverse competenze: IRA (ricevitori e meccanica), OAArcetri (antenna, calibrazione, beamforming), OACatania e IASF-Milano (sistemi di acquisizione), per un totale complessivo di circa 45 FTE investiti negli ultimi 6 anni.

Un investimento in ambito tecnologico da parte di INAF su LOFAR 2.0 è richiesto dall'attuale roadmap (Sez.1.3). L'investimento prevede circa 6 FTE equivalenti nel corso dei primi 3 anni dall'ingresso di INAF in ILT, di cui si prevedono ragionevolmente circa 3 FTE effettivi e circa 150 kEuro di investimento in costi di prototipizzazione e supporto del gruppo. **Questo investimento è pienamente sostenibile dal gruppo LFAA INAF** anche senza prevedere necessariamente l'acquisizione di nuovo personale.

Una collaborazione con ASTRON su LOFAR 2.0 rappresenta un'opportunità significativa per la comunità LFAA INAF ed è uno sviluppo fortemente connesso anche alle attività in itinere su SKA-LOW potendo quindi sfruttarne le competenze già sviluppate in ambito INAF.

Allo stesso tempo tale investimento permetterebbe alla comunità INAF di acquisire un ruolo importante all'interno di LOFAR 2.0, essendo l'unico partner non Olandese ad avere un ruolo nello sviluppo tecnologico dell'infrastruttura, creando anche condizioni favorevoli per acquisire finanziamenti esterni (es Europei etc).

Collaborazioni fra INAF e ASTRON sono peraltro già esistite e/o in itinere in ambito SKA, ricordiamo ad esempio lo sviluppo della catena di ricevitori da parte di INAF nell'ambito del progetto EMBRACE nonché gli stessi programmi AAVP e AADC.

In ambito LOFAR il gruppo LFAA INAF ha portato avanti una campagna di caratterizzazione e calibrazione della Stazione LOFAR CS302 in Olanda tramite un complicato sistema basato su voli di drone nell'ambito del progetto PRIN TECNO 2014 "*Advanced calibration techniques for next generation low-frequency radio astronomical arrays*".

Da Ottobre 2017 sono stati portati avanti dei negoziati esplorativi con ASTRON per definire una roadmap tecnologica in ambito LOFAR 2.0. Le condizioni al contorno di tale roadmap sono state fissate dal Board di ILT che aveva suggerito espressamente alle 2 parti di definire una collaborazione che si concentrasse su linee essenziali e con aspetti di criticità (es attuale mancanza di manpower o di competenze specifiche in ASTRON) del progetto LOFAR 2.0. Allo stesso tempo la richiesta di INAF nel corso dei negoziati è stata quella di concentrare il proprio investimento tecnologico su linee tecnologiche di interesse per l'Ente e sulle quali il gruppo LFAA o parte di esso potesse rivestire ruoli di PI-ship avendo sia un ritorno di visibilità che di know-how.

Le discussioni sono ancora in itinere ed è naturale che un certo numero di dettagli potranno essere definiti solo nei prossimi anni. **Tuttavia le parti hanno già prodotto un testo di accordo programmatico identificando le linee di collaborazione chiave, le modalità della collaborazione e i prodotti attesi (Allegato 5).**

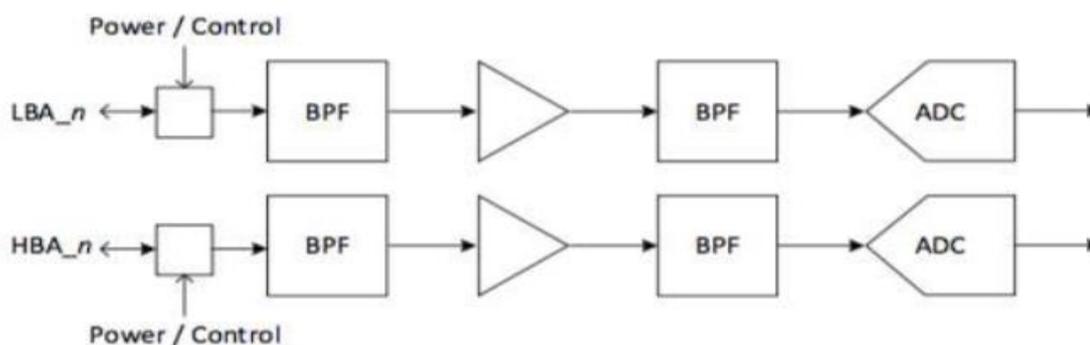
Di seguito riassumiamo le linee principali della collaborazione :

1. **Catena di Ricezione RCU analogica** : è la principale linea di collaborazione ed inizierebbe subito dopo l'estate del 2018. INAF collaborerebbe ai requisiti generali di sistema e sarebbe responsabile del design, simulazione e prototipizzazione della parte analogica della catena di ricezione di LOFAR 2.0. Questa parte ha forti aderenze di tipo tecnologico con quanto già fatto da INAF in altri progetti di array a bassa frequenza (EMBRACE, EDA2.0, SAD) ed è perfettamente nelle "corde" del team tecnico di Medicina che si occuperà del design suddetto.

Di seguito si riportano i requisiti funzionali e non funzionali che questa parte dovrà essere "compliance":

Reference	Description
Specification document Andre/Koos	Equipment for LOFAR2.0 shall be backwards compatible with the current LOFAR equipment. (It must be possible to place the subband from the new beamformer at the same center frequency and with the same bandwidth.)
Reference	Description
RD.1	LOFAR2.0 shall observe simultaneously with LBA and HBA antennas.
RD.1	For LOFAR2.0 it is desired to design the equipment such that an extension is possible to observe the complete HBA band (120 - 280 MHz?)
RD.1	Instantaneous RF input bandwidth $\geq 200$ MHz (100MHz low and 100MHz high)
RD.2	SFDR, $>80$ dBc
RD.2	ENOB, $>11$ bit
	Serial output interface (JESD204B)
RD.1	Synchronisation of the ADCs $<1$ ns
Reference	Description
RD.1	Availability of components between 2018 and 2030
RD.1	Cost, the goal is a RCU2.0 cost $< €200,-$
RD.1	Low power, the goal is a RCU2.0 $<5.3$ W
RD.1	Working at an ambient temperature of $50^{\circ}\text{C}$
	Small footprint

In pratica si tratta dello sviluppo di due catene riceventi LB (10-90MHz) e HB (110-230/280MHz) che dovranno essere campionate contemporaneamente dai sistemi di acquisizione (in questo momento la baseline prevede 2 ADC, uno per catena) secondo lo schema in FIG.9..



**FIG.9:** la doppia catena di ricezione per la LB e HB di LOFAR 2.0. vista aerea della Stazione radioastronomica di Medicina.

Le due catene riceventi permetteranno alle antenne LBA e HBA di acquisire segnale in contemporanea permettendo un'osservazione simultanea su tutta la banda, da 10 MHz fino a 230 MHz, e di effettuare la calibrazione della banda LB usando i segnali in banda HB come punto di partenza. Per raggiungere questo obiettivo bisognerà fare un'analisi architetturale ad alto livello e con i dati di input forniti da ASTRON (come il livello delle RFI e le interfacce di IN/OUT da rispettare) si potrà stilare una tabella di specifiche del ricevitore come risultato dell'analisi del sistema. Questa è la base

necessaria per partire con il design vero e proprio del ricevitore e iniziare con le simulazioni di tutte le parti del sistema. Se il risultato delle simulazioni sarà soddisfacente, si potrà pensare di realizzare i primi prototipi, per poi testarli e verificare le prestazioni attese. Il design dovrà tener conto di vari aspetti e dovrà raggiungere il goal in termini di performances RF, ma anche i vincoli meccanici di compatibilità per l'alloggiamento delle schede stesse all'interno del cabinet di LOFAR 1.0. Per cui un design accurato meccanico, in parallelo a quello RF, sarà un'attività che sicuramente dovrà essere condotta. Se tutti i requisiti dei prototipi saranno soddisfatti, l'ultima fase del design sarà l'integrazione dell'RCU analogica con la parte digitale e di acquisizione (in carico ai colleghi di ASTRON), per poi passare infine alla produzione in serie per l'installazione su LOFAR 2.0. **La proprietà intellettuale dei prodotti del progetto sarà condivisa fra ASTRON e INAF.**

Una stima dell'investimento necessario per questa prima linea di collaborazione può essere stimato sulla base dell'esperienza nell'ambito del progetto EMBRACE. Si prevede un lavoro della durata di 1-1.5 anni con un impegno di complessivi 1.6-1.8 FTE effettivi; questi FTE sono già disponibili all'interno del gruppo di Medicina. In aggiunta si prevede un investimento da parte di INAF per costi di prototipizzazione e per il supporto del gruppo (missioni), tali costi sono riassunti nella TABELLA seguente:

	# People	# travels	# Year	Cost each	TOTAL
Mission to ASTRON <1WK	3	5	3	2.000 €	30.000 €
Mission to ASTRON >1WK	2	2	3	3.000 €	12.000 €
Mission in Italy	3	5	3	500 €	7.500 €
					30.000 €

Production of the prototype (it includes production of the board, components, assembling, mechanics)	Cost each	# iteration	
	5.000 €	3	15.000 €

Instrumentation for lab (purchasing+rent)	65.000 €
-------------------------------------------	----------

PARTIAL	110.000 €
Overhead (shipping,new PC/SW/managing) 10%	11.000 €

<b>GRAN TOTAL</b>	<b>121.000 €</b>
-------------------	------------------

Considerando un costo FTE di 66 kEuro l'investimento complessivo da parte di INAF su questa linea di ricerca è di circa 240 kEuro corrispondente a circa il 60% dell'investimento complessivo richiesto da ILT.

2. **Antenne LBA** : al momento non è ancora chiaro il tipo di upgrade che verrà portato alle antenne LBA, potrebbe essere solo ristretto ad alcune componenti dell'elettronica oppure coinvolgere anche l'intera architettura delle antenne. INAF collaborerà allo studio elettromagnetico e della componente elettronica delle antenne LBA per definire possibili upgrade per LOFAR 2.0. Anche questa è una linea di ricerca perfettamente "nelle corde" del team tecnico LFAA di INAF, in particolare del gruppo di Arcetri e di Medicina. Ad esempio alcuni aspetti di criticità delle antenne LBA 1.0 sono state evidenziate proprio da una campagna di misura in collaborazione con INAF usando sofisticate tecniche che sfruttano anche voli di droni (IEEE Transactions on Antenna & Propagation "Strong Mutual-Coupling on LOFAR: Modeling and In-Situ Validation"). Al momento il package sulle antenne LBA di LOFAR 2.0 non è ancora definito completamente e i dettagli si capiranno solo nel corso del prossimo anno. Tuttavia, indipendentemente da questi dettagli, il ruolo di INAF su questa linea comprenderebbe

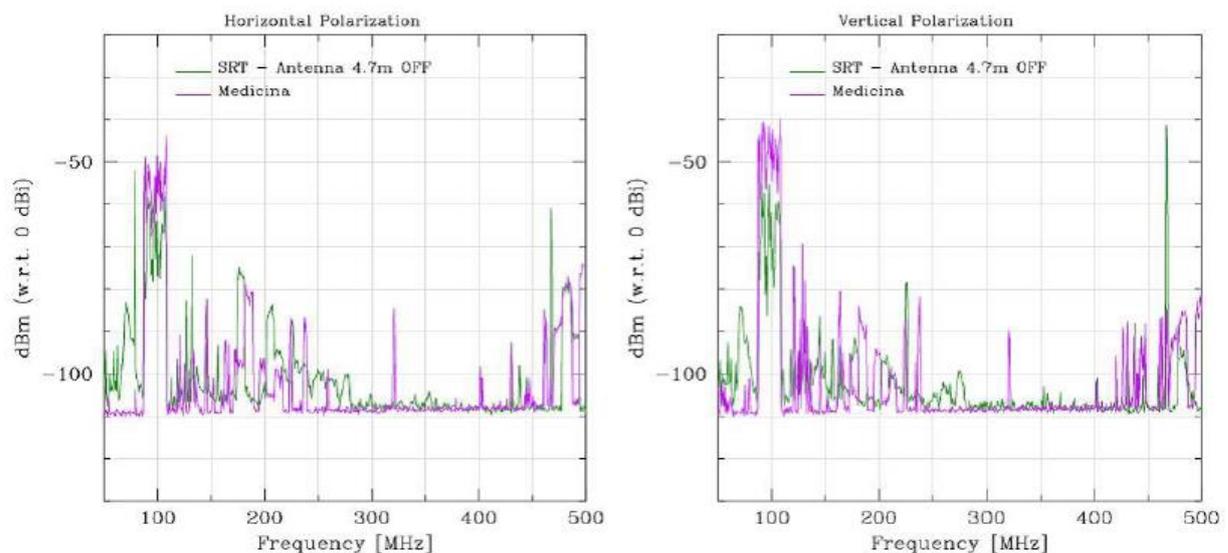
simulazioni elettromagnetiche e campagne di calibrazione che coinvolgerebbero principalmente i gruppi di Arcetri e Medicina.

## ASPETTI RELATIVI AL SITO DI MEDICINA

**Medicina (BO) rappresenta la migliore scelta per ospitare una grande infrastruttura come una Stazione LOFAR 2.0.**

Le motivazioni appaiono molto chiare :

- **STAZIONE DI MEDICINA:** la stazione è una delle 3 sedi di IRA, si trova a circa 30 km da Bologna, si estende su circa 30 ettari di terreno pianeggiante e facilmente raggiungibile. Attualmente ospita la Croce del Nord (collaborazione UniBo e IRA-INAF) e l'antenna VLBI da 32 mt. Vi lavorano circa 20 ricercatori (tecnologi e tecnici), ospita una foresteria e il centro visite Marcello Ceccarelli (<http://www.ira.inaf.it/Home.html>). E' una Stazione radioastronomica radio protetta a cui afferisce anche un gruppo specializzato per la ricerca delle interferenze. La storica Croce del Nord (costruito negli anni 60) è stato uno dei maggiori radiotelescopi a bassa frequenza del cielo nord, operativo alla frequenza di 408 MHz è ancora utilizzato principalmente per la ricerca di space debris. La situazione delle interferenze a basse e bassissime frequenze nell'area della Stazione è monitorata e diverse campagne hanno già dimostrato l'operatività del sito nelle bande LB e HB di LOFAR.



**FIG.10:** distribuzione delle RFI nella banda 50-500 MHz misurata in polarizzazione orizzontale e verticale con antenne LOG periodiche presso il sito di SRT e confrontata con la stessa distribuzione misurata presso l'area della stazione di Medicina (credit Govoni et al 2013, INAF OAC internal Report n.24).

La FIGURA 10 ad esempio mostra un confronto della situazione RFI fra Medicina ed il sito di SRT in Sardegna nel range di frequenze 50-500 MHz. Medicina mostra una banda 50-80 MHz (parte di LB) completamente libera e una situazione accettabile nella banda HB (120-230 MHz); si rammenta che le spikes di interferenze sono rimosse dai sistemi di acquisizione di LOFAR.

- **BASELINE OTTIMALE:** una Stazione LOFAR a Medicina sarebbe importantissima per LOFAR perché permetterebbe di espandere la copertura del piano uv dell'interferometro verso sud estendendo la massima baseline dell'interferometro in direzione NS a circa 1500 km in misura paragonabile all'attuale baseline EW di LOFAR (vedi FIG 7 in Sez. 2). La posizione geografica di Medicina permetterebbe inoltre un

gran numero di ridondanze sul piano uv su baselines di 500-1000 km con la Stazione di Nancay e le Stazioni tedesche. Una volta installata la Stazione a Medicina, ulteriori Stazioni potrebbero essere installate in Piemonte (per ottimizzare la copertura della baseline intermedia fra Medicina, Nancay e Garching) o in Sardegna o Sicilia (soprattutto se la Spagna dovesse aderire ad ILT installando una o più Stazioni LOFAR 2.0).

- **TERRENO:** Medicina offre una soluzione ottimale essendo facilmente ricavabile uno spazio dell'ordine di 1-2 ettari per ospitare la Stazione nei dintorni delle attuali infrastrutture radio. Nell'ambito della LoI per i Premiali 2016 (Allegato 2) il gruppo proponente (che comprende anche il personale afferente al WG) ha anche ispezionato il terreno e provveduto ad individuare un'area (FIG 8, Sez. 2) dove posizionare una Stazione LOFAR 2.0 che per conformazione e per la presenza di un accesso minimizzerebbe anche i costi di preparazione del terreno, stimabili in circa 60 kEuro (Sez.5).
- **STRUTTURA E PERSONALE:** alla Stazione radioastronomica di Medicina afferiscono circa 20 fra ricercatori e tecnici di cui una decina afferenti al gruppo LFAA. Parte di tale personale sarebbe già coinvolto nella collaborazione tecnologica su LOFAR 2.0, assicurando in loco l'FTE necessario per la messa in operatività, gestione e manutenzione della Stazione LOFAR 2.0.
- **CONNESSIONE DATI:** Attualmente la fibra che collega Medicina alla rete GARR ospita due lambda, una lambda "analogica" per gli esperimenti di tempo/frequenza con INRIM ed una lambda da 10 Gbps per i dati. La connessione dati è attualmente utilizzata per flussi da 2 Gbps della durata di alcuni giorni che si ripetono con cadenze mensili; anche considerando un futuro upgrade dei ricevitori è difficile pensare che nei prossimi anni si possa avere un traffico VLBI superiore ai 3-4 Gbps e quindi l'attuale predisposizione della fibra ottica che collega Medicina al GARR è tecnicamente adeguata per supportare contemporaneamente i flussi delle osservazioni VLBI e LOFAR.

Nel caso della stazione LOFAR bisogna considerare i costi dello switch a 10 Gbps che si aggira tra i 5 ed i 10 KEuro. Qualora il traffico diventasse continuativo e non più sporadico come è attualmente nelle osservazioni VLBI bisogna considerare un possibile aumento dei costi dell'ordine dei 20-30 KEuro anno per il trasporto dei dati sulle reti GARR e GEANT verso l'Olanda (vedi Sez.5).