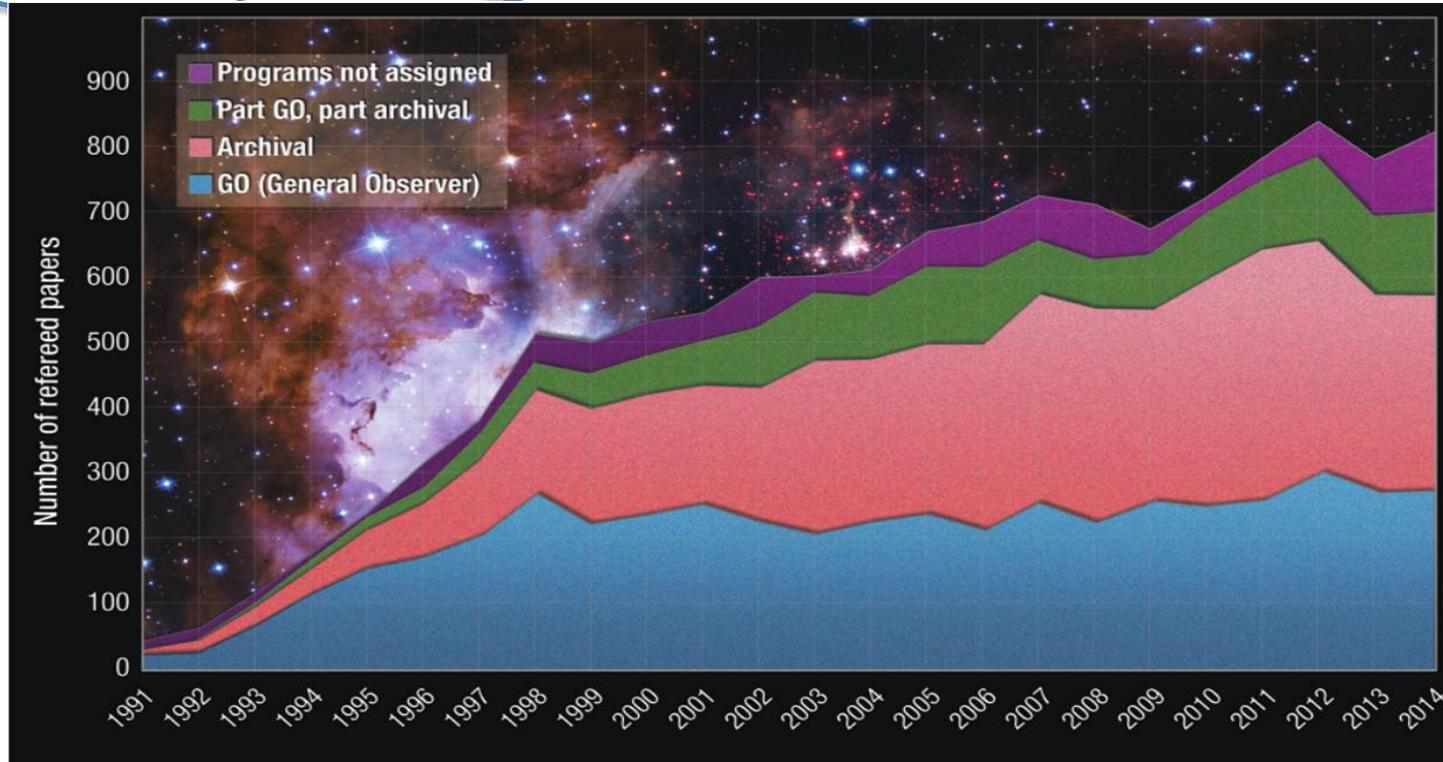


L'evoluzione dell'ICT in INAF: attività e collaborazioni

R. Smareglia

Workshop INFN - CCR – 22-26 Maggio 2017

Archive importance

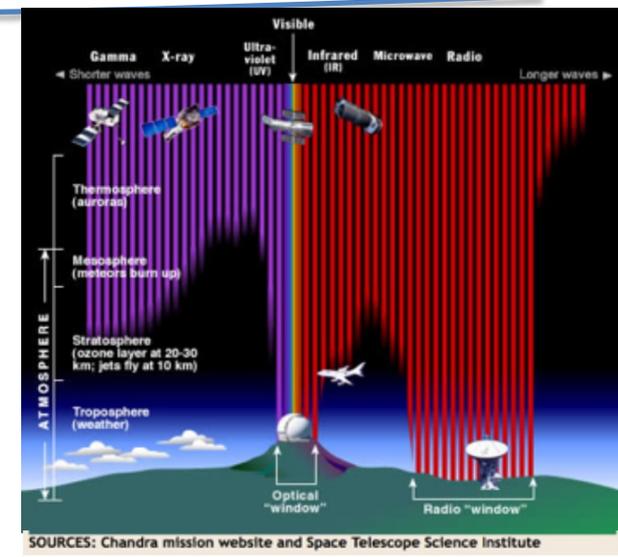


HST Newsletter: “At the present time, approximately **half of the refereed publications** based on Hubble observations are derived purely **from archival data**, and, every year, this number is slightly higher than the number of publications based on new observations. the Hubble Archive has become a goldmine for the astronomical community....”

Lot of work to do



A partial view of the astronomical observatories of the next 10 years



What is missing:

- ▶ some observatories are missing
- ▶ no planetary missions
- ▶ no exoplanets and Sun observatories.
- ▶ no cosmic rays experiments

Main requirement: Interoperability

Distributed resources

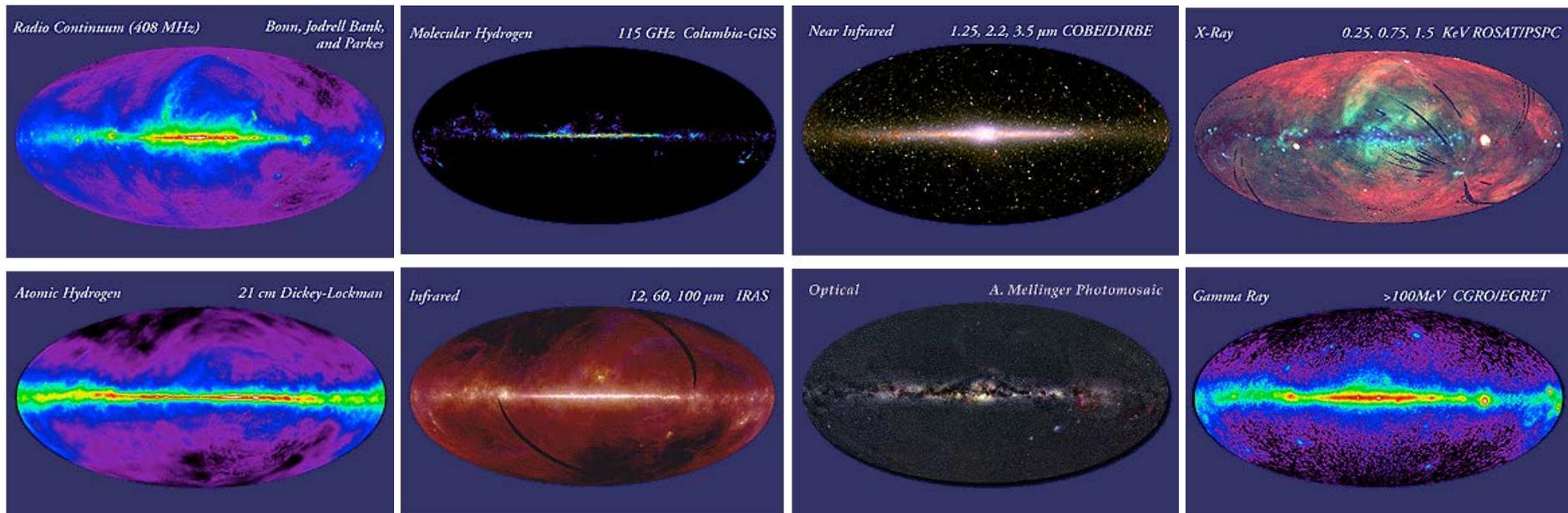
- International team members can bring regional resources
- Big data: moving code to data
- Resources are not simple

Science teams

- Science teams are international virtual organisations
 - Forming around a given multi-year project
 - Handling large datasets
 - Faced with acquiring and building project infrastructure
- Require infrastructure
 - Larger datasets
 - Data management, data distribution, data processing
 - Challenging a team's ability to produce and maintain infrastructure
- May have access to national and regional infrastructure

Courtesy of S. Gaudet

A multi-wavelength digital sky that can be searched, visualized, and analyzed in new and innovative ways
Space, Ground, and Theory data



Courtesy of P. Fabiano

What is the Virtual Observatory?



The VO is a paradigm for Supporting interdisciplinary and collaborative research in astronomy and exploiting the full power of growing and emerging data sets

The latest stage of good data practices in astronomy

FITS provided a first standardization, the Virtual Observatory is the natural progression towards interoperability of data, services and tools

The VO is a framework

- For data centers to provide co-operating data services,
- For software providers to offer a variety of compatible analysis and visualization tools and user interfaces

Interoperability: Open Access - Open Science

Open Access and Open Science is one of the MUST of the EU/H2020 funding project policy

- *The **European Open Science Cloud** (EOSC) pilot project, in which INAF is involved, will support the first phase in the development as described in the EC Communication on European Cloud Initiatives [2016].*
 - *It will establish the governance framework for the EOSC and contribute to the development of European open science policy and best practice;*
 - *It will develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains; and*
 - *It will engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research*

Roadmap per calcolo INAF-INFN

Area di riferimento:

- Progetti sviluppati internamente ai due Enti che prevedono attività di simulazione, data analysis e archiviazione dati.
- Progetti nazionali ed internazionali già in corso e comuni ai due Enti e che richiedono risorse di calcolo per attività di simulazione, data analysis e archiviazione dati.
- Progetti nazionali ed internazionali in corso e comuni ai due Enti e che richiedono risorse di calcolo per attività di simulazione, data analysis e archiviazione dati.
 importante per la partecipazione congiunta ai progetti finalizzato ad
- Progetti nazionali ed internazionali in corso e comuni ai due Enti e che richiedono risorse di calcolo per attività di simulazione, data analysis e archiviazione dati.
 importante per la partecipazione congiunta ai progetti finalizzato ad

If you want to run fast, run alone; if you want to run far, run together.

- an African proverb

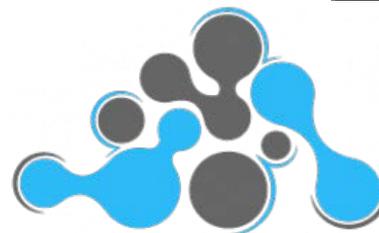
Obiettivi di un Accordo Quadro.

- valorizzazione delle risorse dei due Enti in un contesto condiviso
- valorizzazione delle competenze e della complementarietà
- condivisione delle risorse salvaguardando le specificità ma evitando inutili duplicazioni
- partecipazione congiunta a call Europee di interesse comune

Common Projects



INDIGO - DataCloud



EOSC pilot

The European Open Science
Cloud for Research Pilot Project



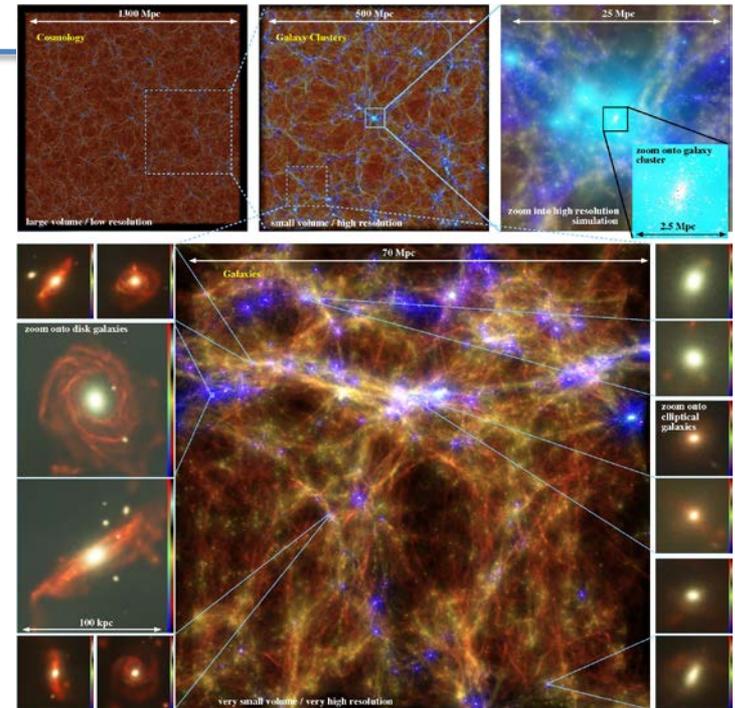
Astronomy ESFRI & Research Infrastructure Cluster



- INAF had more than 90 software packages developed, some public, many “locally” to be engineered
- Raw data is public, but “science ready” data is not yet.
 - ➔ Using DOI to suggest share experience and work (software, data, gray articles, ..)

- Non e' piu' solo una questione di simulazioni teoriche ma anche di analisi dati ...
- Software ed expertise “regalati” all'esterno per mancanza di infrastrutture
(Perdita di “paternita’” – co-authors)
- Mancanza di informazione su cosa ci potrebbe essere e su chi sono gli esperti
(the answer is .. Lost .. In the wind)

Ex.: Simulation HTC/Cloud



- Personale INAF coinvolto:
 - Staff ~200
 - Contrattisti/associati ~300
- Cores complessivi in uso oggi: ~70000/day
- Storage complessivo: ~1200TB

Non esiste “il modello di calcolo”

- Piu' livelli
 - 8 \Leftrightarrow 64 Core (“in casa” ... forse)
 - 64 \Leftrightarrow 1024 Core (Tier X.y (2/2.5))
 - Pochi Core tanta Ram
 - Tanti Core poca Ram
 - > 1024/2048 Core Tier 0 @ CINECA
 - GPU
 - Acceleratori
 - ...

➤ Tier 0 → CINECA (Mou signed)

➤ E-Infrastructura INAF :

- Calcolo → CHIPP Project
 - HPC
 - Cloud / Virtualizzazione
 - GPU
 - FPGA
 - ...

➤ Archivi

➤ Rete

➤ Fornire Servizi

- HW
- Personale
- Help desk
- Netwoking



Interno e/o outsourcing

Quindi...

- Sistemi Monolitici non sono la soluzione
- Nessuno ti regala niente
 - ➔ le partnership hanno pro/contro
- Domanda e' cosa vuole INAF:
 - Creare una cultura per essere pronti tra 5-10 anni
 - Demandare completamente ad "altri"

Partnership: OpenPOWER foundation.....

- Investments and research projects for the availability and better exploitation of the new HPC generation is a priority for INAF to maintain excellence level in many key projects, and to maintain the competitively
- **INAF** becomes a member of the **OpenPOWER Foundation** in 2015 (Academic member)

- The OpenPOWER member Foundat customer Linux) a
- This a IBM.
- These cente optin

Implementation / HPC / Research

Software

System / Integration

I/O / Storage / Acceleration

Boards / Systems

Chip / SOC

s. The
ducer-
d on

s on
data
rm
ation.

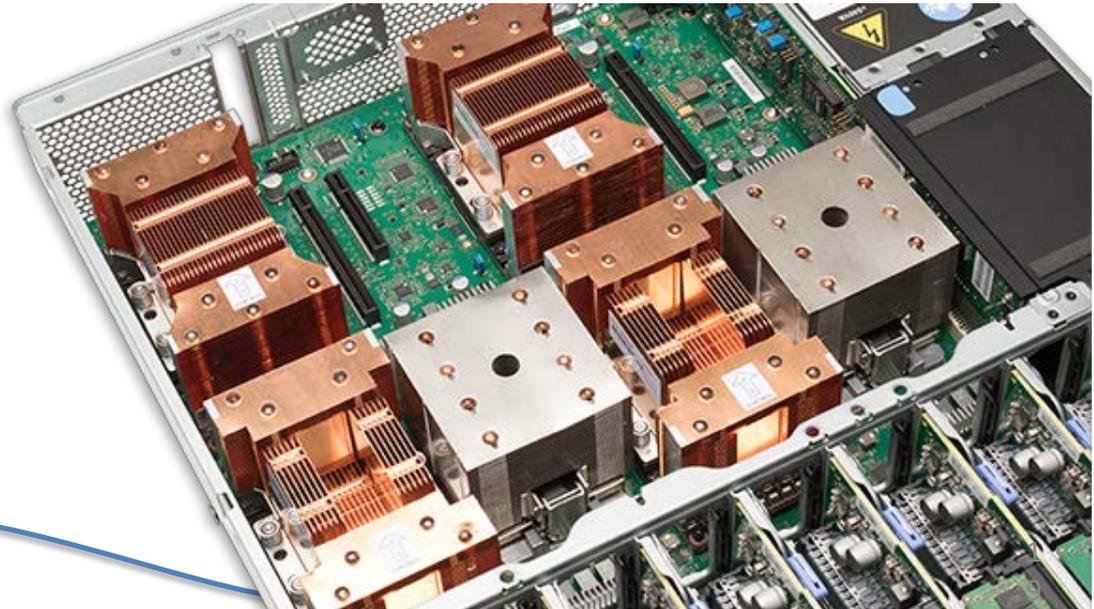
OPF FOR PHYSICAL SCIENCE

WG: PURPOSE

- INAF have proposed a workgroup for Physical Science needs within the OpenPOWER Foundation:
OpenPOWER Foundation for Physical Science Workgroup
 - The proposed workgroup aims at addressing the challenges of **Physical Science projects**.
- The main **purpose** is to have a forum of 'scientists' and 'technological developers' at the same level around a technological solution (Power architecture and Linux). Some advantages:
 - a direct connection with hw/sw developers
 - a direct and different connection with the market
 - To understand where technology is going and help address it as an active part, and not just a “consumer”

IBM S822 LC for HPC

- @INAF we have a Power8+ machine for test and development
 - S822LC for High Performance Computing
 - Two 8-core 3.25 GHz POWER8 processor cards (128 threads)
 - 2 GPU P100
 - 512 GB memory
 - 2 TB SSD HD
- We are testing the machine for Deep Learning, Monte Carlo simulations and porting of code.



- ✓ All INAF structure
 - ✓ About 54 arc
 - ✓ 59% pub
 - ✓ **Policy IN**
- ✓ Italian center for
- ✓ Alma Regional Ce
- ✓ GAIA (on-fly) → [
 - ✓ 1 PB (mainly
- ✓ Euclid → > 10 x C
- ✓ CTA (ASTRI) → > 10 TB/giorno
- ✓ SKA → > 100 TB/giorno

ONLINE MANUSCRIPTS @ VATLIB.IT



- ✓ Data Curation & Preservation
 - ✓ Standard FITS (from 1970)
- ✓ Data Interoperability → Virtual Observatory



IA2 Italian Center for Astronomical Archives
Centro Italiano Archivi Astronomici



- Home
- Services ▾
- Projects ▾
- Software
- Additional Info ▾
- IA2 Group

IA2 (Italian center for Astronomical Archive) is an Italian Astrophysical research e-infrastructure project that aims at co-ordinating the quality of astrophysical data services. It aims at co-ordinating these developments and facilitating access to this data supported by INAF since 2005. IA2's main goals are:

- archiving systems and safety;
 - data curation and preservation;
 - distribution over geographical sites;
 - data publication in the VO;
 - hosting data;
 - development of data handling, VO and archive software;
 - providing service data workflow an
- [contact us](#)



Main page

- ▾ About us
 - ARC staff
 - Areas of expertise
 - The ARC cluster
 - Contact us
- News and documents
- Lectures and thesis
- User Support
- ALMA useful links
- Other ARCs
- Private Area
- Tools

IA2 Services

Archives

Managed from IA2

- **TNG** (Telescopio Nazionale Galileo)
 - old Archive
- **LBT** (Large Binocular Telescope)
- **Asiago**
- **Serra la Nave**
- **Medicina/Noto Radio Archive**

Hosted

- **Bjurakan**
- **INES**
- **Basti**
- **TIRGO**

Virtual Observatory

- SIA
- TAP

Data Sharing & Data Preservation

Owncloud Data Preservation

More services

Workflow (Yabi) DOI

IA2 Projects

EU Funded project:

FP7: VIALACTEA GENIUS
H2020: ASTERICS INDIGO AENEAS

Other Project:

GAPS Prisma

IA2 Software

Page Discussion

Read View source View history

Log in

Italian ALMA Regional Centre



EUROPEAN ARC
ALMA Regional Centre || Italian

The Italian node of the European ALMA Regional Centre is hosted by the Istituto di Radioastronomia in Bologna and is one of the seven nodes that constitute the European network that provides technical and scientific support to ALMA users. The nodes operate in close collaboration with each other and with the ALMA Regional Centre at ESO, Garching. Each node contributes its own specific expertise, in order to ensure that maximum advantage is taken of the European competences in the field of mm-astronomy and interferometry.

Our ARC node staff support the ALMA users in all the steps of their projects, by helping in

- ALMA proposal preparation and submission
- optimising the observing strategy
- tracking the project status
- reducing interferometric data with CASA
- archive mining
- handling large data sets
- polarimetry
- mm-VLBI with ALMA

Contents [hide]

- 1 Five years of Early Science with ALMA and the Italian ARC-node
- 2 Next events at the Italian ARC:
- 3 Past Events at the Italian ARC:
- 4 ALMA News:

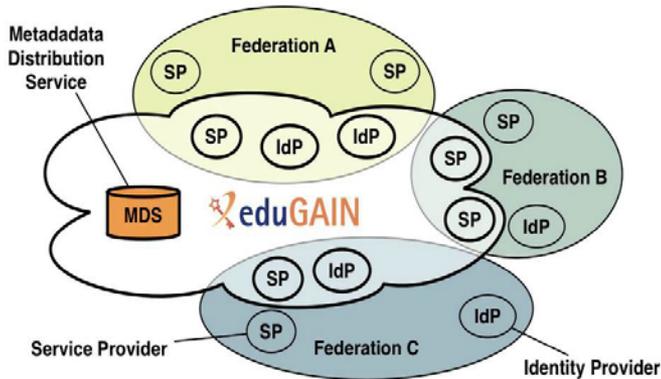
Five years of Early Science with ALMA and the Italian ARC-node

Five Years of Early Science: a look at the performance of the Italian community, and the activities of the Italian ARC-node

Next events at the Italian ARC:



Authentication (SAML based)



Authorization Internet2 application based / VO compliant based



Future plans of interoperability:

What about RAP????

- IA2 IdP is based on a replica of the Bologna LDAP that provide Digital Identities to INAF (IDEM);
- IA2 integrate the LDAP with non IDEM identities that use to one or more IA2 Telescopes;

Remote Authentication Portal

Image Credit & Copyright: Colombari/E. Recurt

| | | |
|---|---|--|
| Use the eduGAIN Logo to Login or Register to the RAP facility if you belong to an eduGAIN | Use the Google Logo to Login or Register to the RAP facility with your social identity. | Use the X.509 Logo to Login with your personal certificate (TERENA, GARR and INFN CA are |
| Use the Local Logo to Login with your self registered account. | Login with your Username and the received RAP Token, if your remote providers is unreachable. | Remote Authentication Portal was written by Franco Tinarelli at INAF-IRA |

SKA Regional Centres



SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Courtesy of A. Chrysostomou

One Observatory: The Square Kilometer Array

Two Telescopes

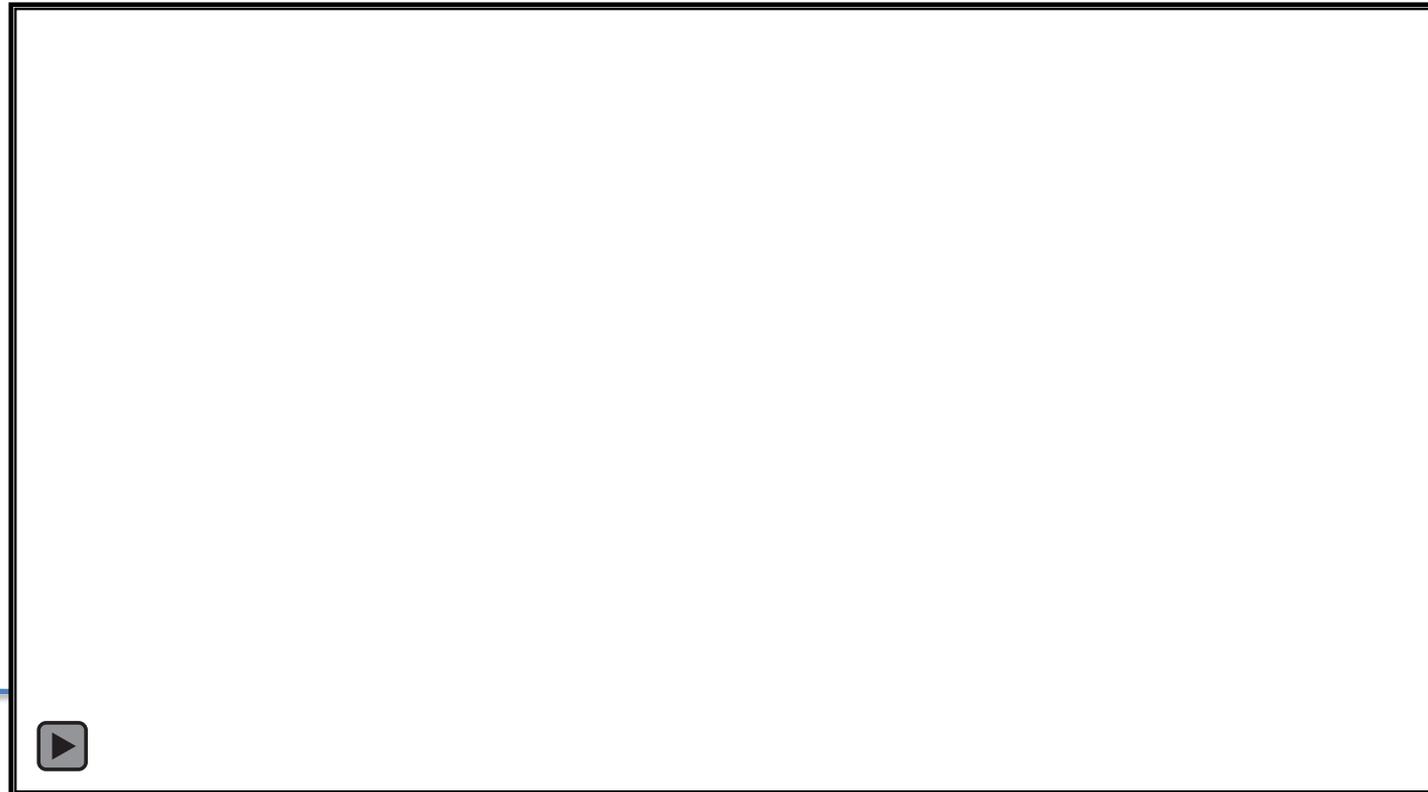
- SKA-LOW
- SKA-MID

Three Sites

- Australia (LOW)
- South Africa (MID)
- UK (GHQ)

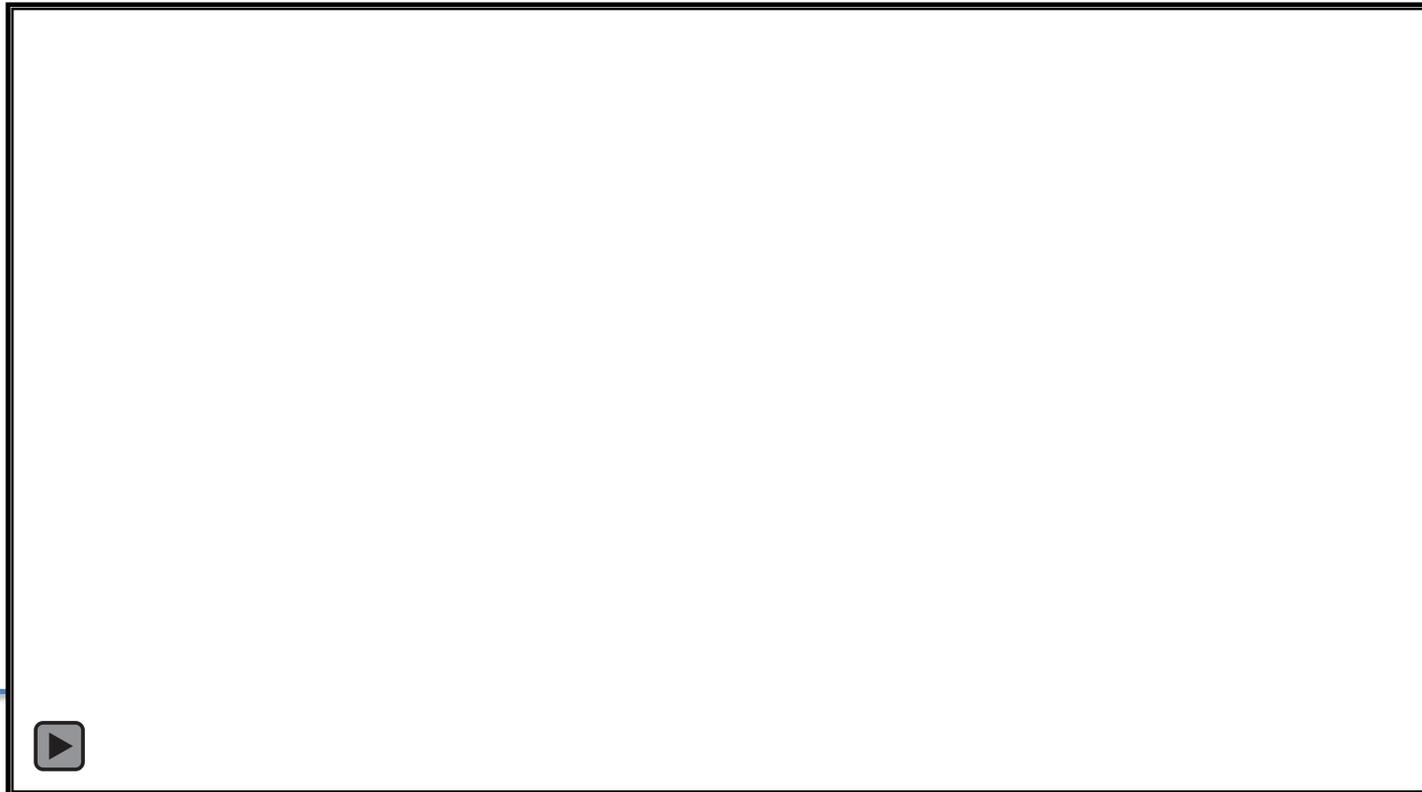
SKA1 LOW - Western Australia

- 131,072 antennas : 512 stations of 256 antennas, core + 3 spiral arms, 65km baselines
- 50 → 350 MHz full instantaneous bandwidth
- Raw Data output approx. 2 Pbit/s → 7 Tbit/s into the correlator



SKA1 MID - Karoo, South Africa

- 133 SKA1 dishes (15m), 64 MeerKAT (13.5m), core + 3 spiral arms, 150km baseline
- 0.35 → 15GHz covered in 5 bands
- Raw Data output approx. 9 Tbit/s into the correlator



Data flow

SKA1-LOW



Photo: Chryssostomou/CSIRO



SKA1-MID

Courtesy of A. Chrysostomou

Some perspective

SKA Science Archive

searches on
Google
98PB

uploads to
facebook
180PB

YouTube
15PB

CERN
15PB

LOFAR
Long Term Archive
23PB

ESA
6PB

Census
4PB

HARDAG
3PB

LIBRARY OF CONGRESS
5PB

SKA
Phase1 Science Archive

300PB

PER YEAR

● 1 Petabyte

A collaborative model of SRCs (SKA Regional Center/s)

Three main factors that lead to a model of a **collaborative network** of SRCs:

- (1) The science data products that emerge from the SKA observatory are not in the final state required for science analysis
- (2) The data volumes are so large that direct delivery to end users is unfeasible
- (3) The community of scientists working on SKA science data products will be geographically distributed

Grazie per l'attenzione

