2019 ICT workshop

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

Istituto di Astrofisica Spaziale e Fisica cosmica di Milano

24 OCTOBER 2019

On the way to the SKA Regional Centres

ANDREA POSSENTI

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

National landscape

HPC

- **CINECA** Italian super-computing centre that offers HPC peta-scale computing facility (~25 PetaFLOPS)
- **SISSA** "Ulysses" cluster with about 7000 Cores for scientific computing and a Master in HPC to train new scientists.
- **ENEA** CRESCO Cluster for about 5000 Cores dedicated to scientific computing
- **INFN** computing Infrastructure

NETWORK

- **GARR** national infrastructure

  Backbone 400 Gbps up to 1 TBps soon
- some **INAF** Structures and Observatories on 10 Gbps at the moment:
  - Antennas VLBI (SRT, Mc, Nt) ⇒ 10 Gbit/s
  - OATrieste ==> 10 Gbit/s
  - Upgrading to 10 Gbit/s: OACAgliari ,IRA Bologna, OACatania

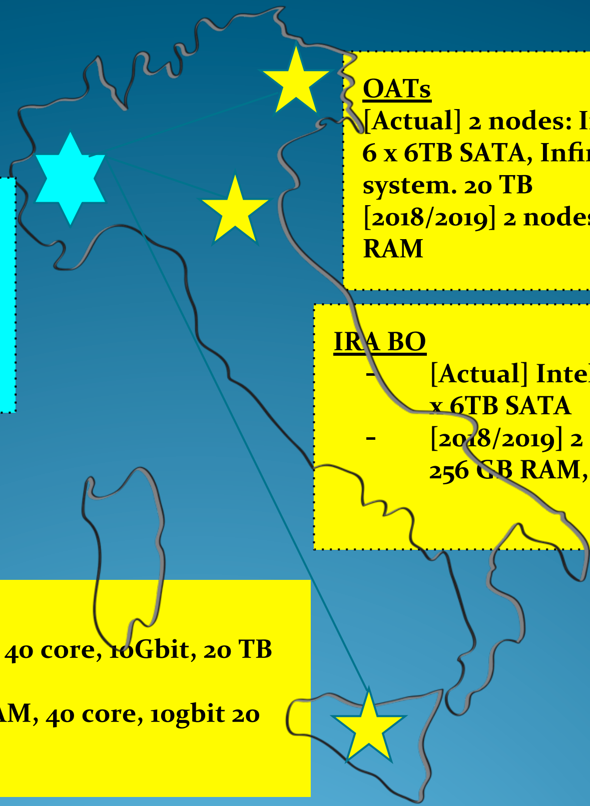
HTC & CLOUD

- **Italian Computing and Data Infrastructure**: includes major c institutions involved in HTC, HPC and Cloud computing**: INAF, INFN, CNR, ENEA CINECA, GARR**
- **GARR Cloud** offers cloud services to the Italian academic and research community based on open standards
- **INFN** distributed computing infrastructure for HTC and Cloud built for High energy physics experiment (LHC) with a main node (TIER-1) in Bologna

INAF computing and data storage resources

- INAF distributed HPC/HTC infrastructure that involves different sites in Italy and offers a computing resource for ~ 25 TFLOPS and HPC storage for about 500 TB
- INAF – CINECA MoU to offer HPC resources for Astronomers
- INAF national service computational resources CHIPP
- Cagliari HTC-HPC system will offer soon 12+ nodes dualCPU-dualGPU
- LOFAR.IT distributed infrastructure based on 4 sites in Italy to offer HTC resources for LOFAR data reduction and analysis
- INAF cloud service offers a EOSC compatible cloud access to computing and storage resources based on OpenStack
- INAF archive/storage
    - IA2 data center (0.5 PB on disk + 8 PB on tape)
    - SRT data center (2.5 PB on disk + 8 PB on tape)
    - More than 3 PB on disk shared across several struct

… an example: LOFAR-IT infrastructure



UNITO
3 FAT node on OCCAM
4 x Intel Xeon (12 core)
RAM 768 GB DDR4, 1 SSD 800GB, 1 HDD 2TB, 2x10Gb

OATs
[Actual] 2 nodes: Intel Xeon, 512 GB RAM DDR3 6 x 6TB SATA, Infiniband ConnectX®-3 - Hotcat system. 20 TB
[2018/2019] 2 nodes: 2 socket (40 core) 256 GB RAM

IRA BO
- **[Actual] Intel Xeon 512 GB RAM DDR3 6 x 6TB SATA**
- **[2018/2019] 2 nodes: 2 socket (40 core) 256 GB RAM, 10gbit ethernet**

OACT
- **[Actual] 2 nodes: 256GB RAM, 40 core, 10Gbit, 20 TB Storage**
- **[2018/2019] 2 nodes: 256 GB RAM, 40 core, 10gbit 20 TB storage Bee-GFS.**

INAF CLOUD.OATS.INAF.IT powered by openstack

Relevant Projects

- **AENEAS**: https://www.aeneas2020.eu/ Italy leads WP5 and has task leaders in many WPs
- **ESCAPE**: https://escape2020.eu/wp_escape.html Italy leads some tasks
- **EOSCPilot**: https://www.eoscpilot.eu/ Cloud Project where INAF participate in main activities and is involved in porting software analysis tools in the cloud (EOSC)
- **Exanest**: http://www.exanest.eu/ European Exascale System Interconnect and Storage
- **EuroExa** https://euroexa.eu/ co-design of innovative exascale system: INAF is involved in the co-design of Exascale infrastructure with the application porting
- **IA2** data center
  - Authentication Authorization and Accounting (AAA) actions
  - First SKA Science Data Challenger
  - User Space and INAF Radio Science Gateway
  - Radio archive
- **SRT** data center
- **ALMA** regional centre
- **VO** activities (IVOA)

Collaborations (active and in progress):

- LOFAR: Pipeline parallelization
- ASKAP: source finding tools
- Meerkat: Virtual reality for structure recognition
- MeerKATHI: pipeline parallelization

The astro-computational quickly evolving world context…

Available estimates suggest that the **SKA Observatory will generate <u>600 PB</u> of calibrated science data products <u>each year</u>**. This data rate is unprecedented in observational astronomy.

The infrastructure for transporting such large data volumes to users around the world, and the computational resources that are required to enable users to turn those data into scientific results, are **NOT** within the current planned scope of the SKA project and demand imaginative solutions.

… some of the challenges …

Sse Mauro Nanni's talk

- ✧ **Huge data volumes to transport**
- ✧ **Unprecedented number of sources per pointing to extract and characterize**
- ✧ **Data visualization**
- ✧ **Calibration**
- ✧ **Polarization Calibration**
- ✧ **RFI excision in presence of very large number of frequency channels**
- ✧ **Data archiving**
- ✧ **…..**

The aims of the Ska Regional Centres (SRCs)

July 2016: the SKA Board deliberated:

"The SKA Observatory will coordinate a network of SKA Regional Centres that will provide the data access, data analysis, data archive and user support interfaces with the user community"

November 2018: the SKA Board deliberated:

"The mission of the SRC Steering Committee (SRC-SC) is to define and create a long-term operational partnership between the SKA Observatory and an ensemble of independently-resourced SKA Regional Centres.

The SRC-SC will be superseded in due course by the operational partnership that is formed as a result of its work"

The responsabilities of the SKA Observatory and of the Ska Regional Centres (SRCs)

The SKA Observatory and the SRCs will be jointly responsible for:

a) maximizing the quality of SKA data delivered to users;
b) the production of Advanced Data Products;
c) ensuring that the approved science program can be accommodated within available resources;
d) ensuring the availability of a platform of distributed services across computational and data infrastructures to support the user community to deliver SKA science.

The SRC-Steering Committee (SRC-SC)

Peter Quinn	Australia
Severin Gaudet	Canada
An Tao	China
Jean-Pierre Vilotte	France
Yogesh Wadadekar	India
Andrea Possenti	Italy
Michiel van Haarlem	The Netherlands
Simon Ratcliffe	South Africa
Lourdes Verdes-Montenegro	Spain
John Conway	Sweden
Anna Scaife	United Kingdom
Antonio Chrysostomou	SKA Organisation
Rosie Bolton (secretary)	

The SRC-Steering Committee (SRC-SC): action plan and tentative timeline

Elaborating a white paper in which a general description of the rules and operations of the SRC-SC will be drafted — Nov 2019

Discussing that with the involved community — End 2019

Start assembling the Working Groups needed for the implementation — Early 2020

Refine and finalize a document to by presented at the SKA Board — Spring 2020

Upon approval, start organizing the implementation of a proto SRC network — Mid 2020

Delivering a working proto SRC network — Sometimes 2021

White paper layout

Sections

- Complete operating descriptions for end-to-end SKAO/SRC use for all user groups
- Common SW platform
- Community engagement roadmap, plan for initiation of user fora
- Description of the SRC body
- White paper needs to identify sizing components, cost/size scaling

Working Groups:
to be established after the Shanghai meeting

A preliminary list of working groups

Users

- **Telescope users**
- **Archive users**
- **SKA Observatory**
- **Developers**
- **Tools users (SW interoperability)**
- **Non-astronomy users**
- **Commercial users**

Operations (SRC + SKAO)

Software including services

Global networking and Data logistics

Archive including IVOA

Prototyping and SRC data challenges

FAIR

Italian expected outcome of the SRC-SC activities

✓ **1. The identification of a kernel of "modi operandi" in the interactions among the various actors to secure an efficient and always developable science-needs driven system**

Examples of initial possible functionalities for a Common Platform

The functionalities of this platform would include:

- **Integration of (heterogeneous) storage resources provided by different facilities in order to offer a common space for storing data (similar to the Data-lake concept of ESCAPE project)**

- **Integration of (heterogeneous) computing resources provided by different facilities**

- **Uniform access to the SRC services irrespective of their geographical location**

- **AAI services for the users, so that they do not require different credentials to access different resources/SRCs**

Examples of initial possible functionalities for a Common Platform

A preliminary list of the services which this platform would require....

- An environment allowing the implementation of workflows / notebooks in a collaborative way

- An execution engine able to capture provenance, and store a workflow each time that is executed to generate Advanced Data Products

- A data archive which includes provenance information gathered from the instrument and the subsequent workflow (item above)

- A service to access the data following VO standards allowing interoperability and multi-wavelength/multi-messenger science

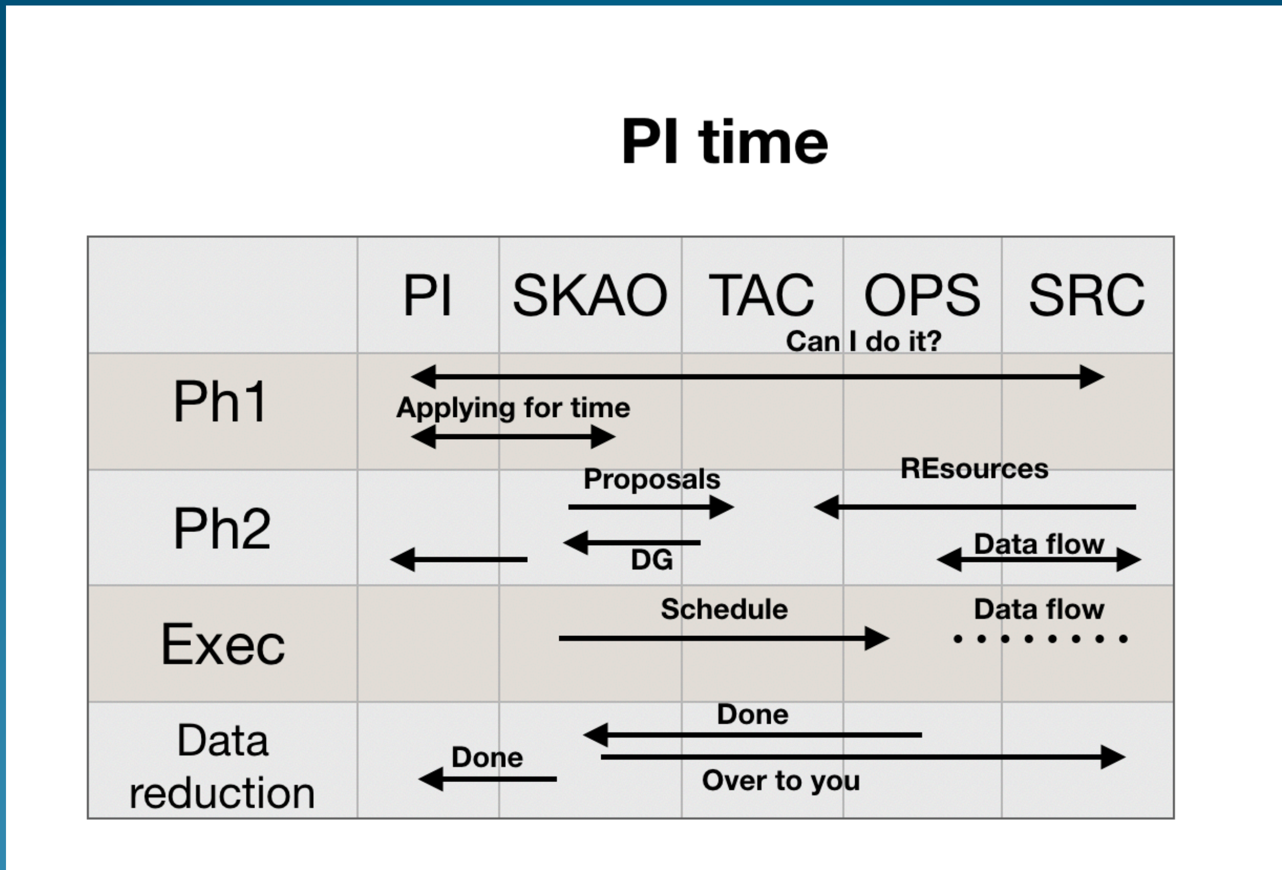
- A catalogue of workflows/pipelines/code that can be customised to facilitate reuse and avoid reinvention

Some key points under discussion

- Governance of worldwide SRC system as a unique entity knowing all resources
- How to ensure SRCs continuity
- Data management will be undertaken by the SRC body
- Data placement in SRCs will be driven by optimising science
- There will be a minimum set of SW enabling a common SW platform
- Allocation process which considers SKAO resources alongside SRC resources
- FAIR principles for data and methods
- National contributions to the SRC pool in terms of capability and not cost
- SRCs will evolve with the engagement in precursors/pathfinders
- Primary need is to develop an SRC community of users

Building an operational model

A SRC-SC task is to develop timeline diagrams for different use cases, for example building on this below



Italian expected outcome of the SRC-SC activities

- ✓ **2. The establishment of a proto-SRC with a significant pole located in Italy**

Italian SRC pole

INAF President, with CDA endorsement (July 2019):

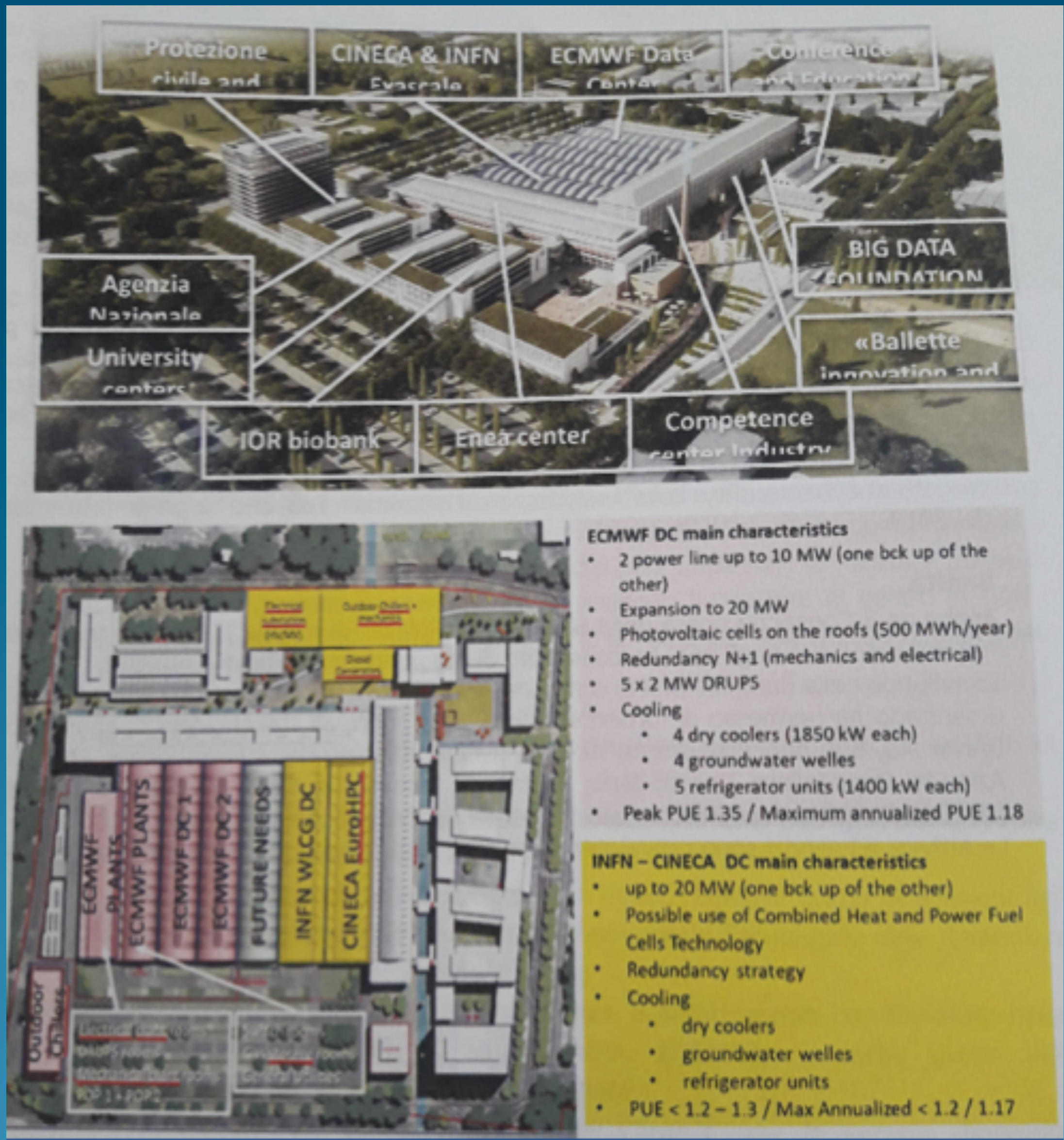
FOR THE INTERNATIONAL PROJECTS CTA AND SKA: CONSOLIDATION OF RESEARCH STREAMS AND DEVELOPMENT OF INNOVATIVE TECHNOLOGIES. ACTIONS FOR INFRASTRUCTURAL CONSOLIDATION ....

" General guidance lines:

1. Investments aimed to the establishment - in the context of the Tecnopolo of Bologna - of a INAF super-centre for IT activities, as a "Regional Center" for the data-processing of SKA and CTA "

Italian SRC pole

The chance of the Tecnopolo in Bologna



Italian SRC pole

The opportunity of the Tecnopolo in Bologna:

Tecnopolo in Bologna will host:
- Leonardo 270 PFlops
- ECMWF
- INFN
- CINECA

INAF is considering a location in the area

100GB/s GARR backbone network

>100000sqm for research data centers

Operational in 2020

Up to 20MW power supplies

Shared cooling and power resources

"Thanks to this infrastructure we can candidate as one of the EU SRC for SKA data product analysis"
(D'Amico, INAF President@ MediaINAF)

A progressive building up of the Italian pole of a proto-Regional Centre …

… **it is a plan having also the needed funding** support

Italian expected outcome of the SRC-SC activities

- ✓ **3. The possibility for the regional communities to obtain access to the system (and keep a role of management/development of that) in a fair proportion to the local investments**

Italian involvement in preparatory activities

science

Many tens of Italian astro-scientists are members of the SKA Science Working Groups!

Fundamental role in order to develop requests and imagine solutions to the USE CASES for the SRCs

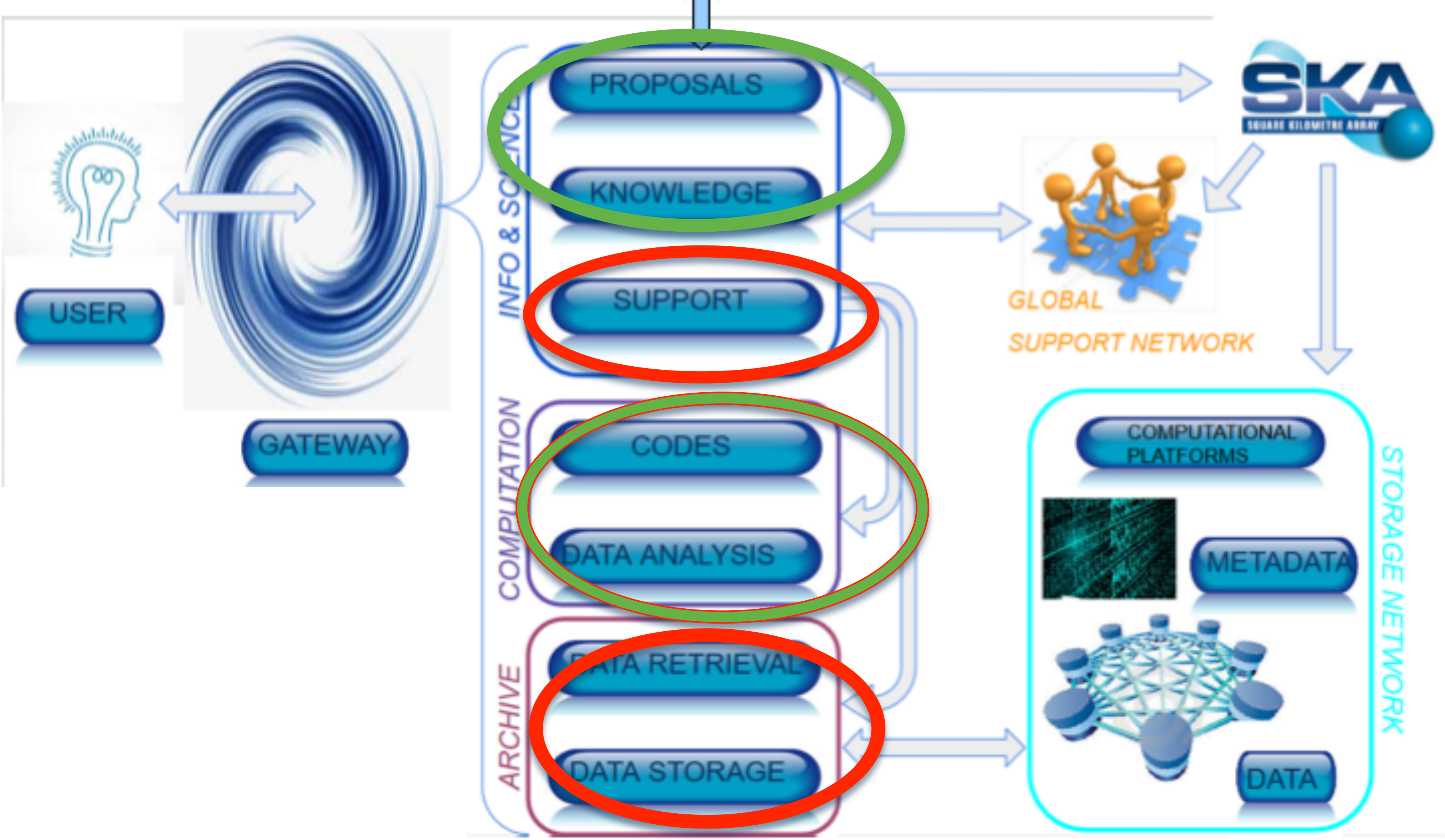
+

Anticipate other foreign astro-scientists in ADAPTING to the new way for doing data reduction an computation in the SKA era

Italian involvement in preparatory activities
software

Instrument	Some key software activities developed within INAF
LOFAR	Data working group Pipelines Computing Code optimization & profiling Porting on exascale machines
ASKAP	Caesar: source extraction & parametrization Algorithms to destripe single-dish images Source extraction from combined IR + Radio
MeerKAT	HI/continuum data analysis for inteferometric data Pulsar pipelines and schedule optimization
uGMRT	Optimization of existing pipelines
eMERLIN	Optimization of existing pipelines – combination with JVLA data
JVLA	Optimization of existing pipelines
Other projects	VisiVo: big astronomical data 3D visualization Distributed data and computer center for SKA

Italian expendable expertise



[edited from Massardi]

Let's exploit these opportunities

THANKS