IN PARTNERSHIP WITH SKAO

The Fourth National Workshop on the SKA Project

26 NOVEMBER 2023 – UNIVERSITÀ DI CATANIA – DIP DI FISICA

SKA Regional Centres (SRC): the development plan

ANDREA POSSENTI

INAF OAC Osservatorio Astronomico di Cagliari

it SRC SKAO Regional Centre Italy

Finanziato dall'Unione europea NextGenerationEU

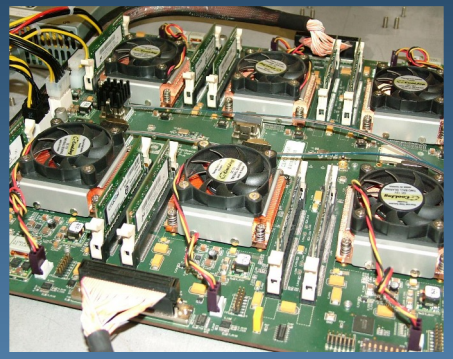
Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

USC VIII

The SKAO data flow

CSP: Central Signal Processor



e.g. FPGAs in the ASKAP correlator

5 Tb/s
data buffer of 2 weeks

SDP: Science Data Processor

e.g. SDP prototype, Cambridge

5 + 9 Tb/s
data
buffer of
2 minutes

600 PB/yr
data persistence

SRC: SKA Regional Centre network

SRC
SRC
SRC
SRC
SRC

SKA portal

Distributed facilities

USERS

Adapted from Philippa Hartley (SKAO)

it | SRC
SKAO Regional Centre Italy

The aim of the Ska Regional Centres (SRCs) and the birth of the SRC-SC

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.**"

it | SRC
SKAO Regional Centre Italy

The SRC-Steering Committee (SRCSC)

Karen Lee-Waddell	Australia
Severin Gaudet	Canada
An Tao	China
Jean-Pierre Vilotte	France
Hans-Rainer Kloeckner	Germany
Yogesh Wadadekar	India
Andrea Possenti	Italy
Akahori Takuya	Japan (observer)
Hyunwoo Kang	Korea (observer)
Michiel van Haarlem	The Netherlands
Tiago Peres	Portugal
Fernando Camilo	South Africa
Lourdes Verdes-Montenegro	Spain
John Conway	Sweden
Emma Tolley	Switzerland
Rob Beswick	United Kingdom
Antonio Chrysostomou	SKA Organisation
Rosie Bolton (secretary)	

it | SRC
SKAO Regional Centre Italy

The statute of the Ska Regional Centres (SRCs)

November 2018: the SKA Board also deliberated:

"The SRC-SC will be superseded in due course by the operational partnership that is formed as a result of its work"

SRC-SC and SKAO have been drafting an agreement during these weeks about the new government, rules and roadmap for the core activities of the SRC network

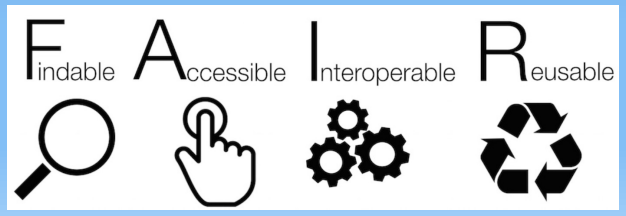
January-February 2024: the Council will review and hopefully finally endorse the plan

it | SRC
SKAO Regional Centre Italy

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

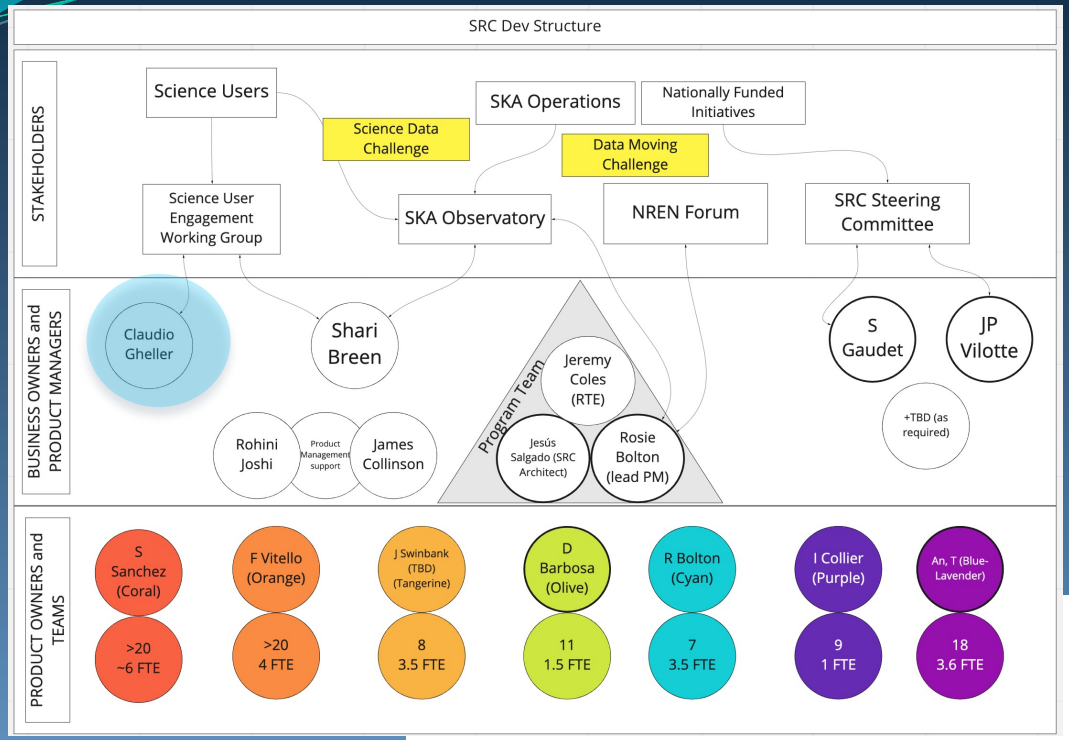
The SKA Observatory and the SRCs will be jointly responsible for:

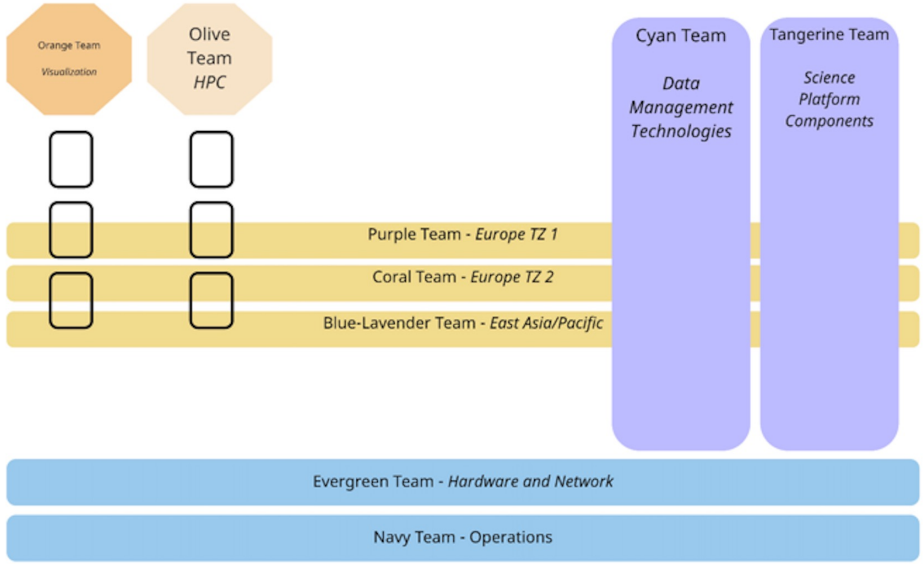
a) maximizing the quality of SKA data delivered to users;
b) the production of Advanced Data Products;
c) storing, archiving and curation of the primary SKA output data and of the Advanced Data Products;
d) ensuring that the approved science program can be accommodated within available resources;
e)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



it | SRC
SKAO Regional Centre Italy

The working structure: the Agile Teams





Enabling teams (experts, multi-time zone, limited in time)

- *Data Management Technologies*: Prototypes 1a and 1b
- *Science Platform Components*: Identify technologies for prototypes 3 and 5

Complicated subsystems teams (experts, multi-zone, in principle not limited in time)

- *Visualization*: Prototype 4
- *HPC*, Cloud, providers

Stream-aligned teams (co-located or single time zone, not limited in time, able to provide value on different areas)

- *Europe Time Zone 1*: Prototype 2 (AAA)
- *Europe Time Zone 2*: Prototype 3 or 5. New features generation: Support to Prototypes 1a and 1b (deployment)
- ***Blue-Lavender Team*** *East Asia + Oceania TZ 1*: Prototypes 3 or 5. Support to Prototypes 1a and 1b (deployment)

it | SRC
SKAO Regional Centre Italy

The ongoing work: The prototypes

Prototypes

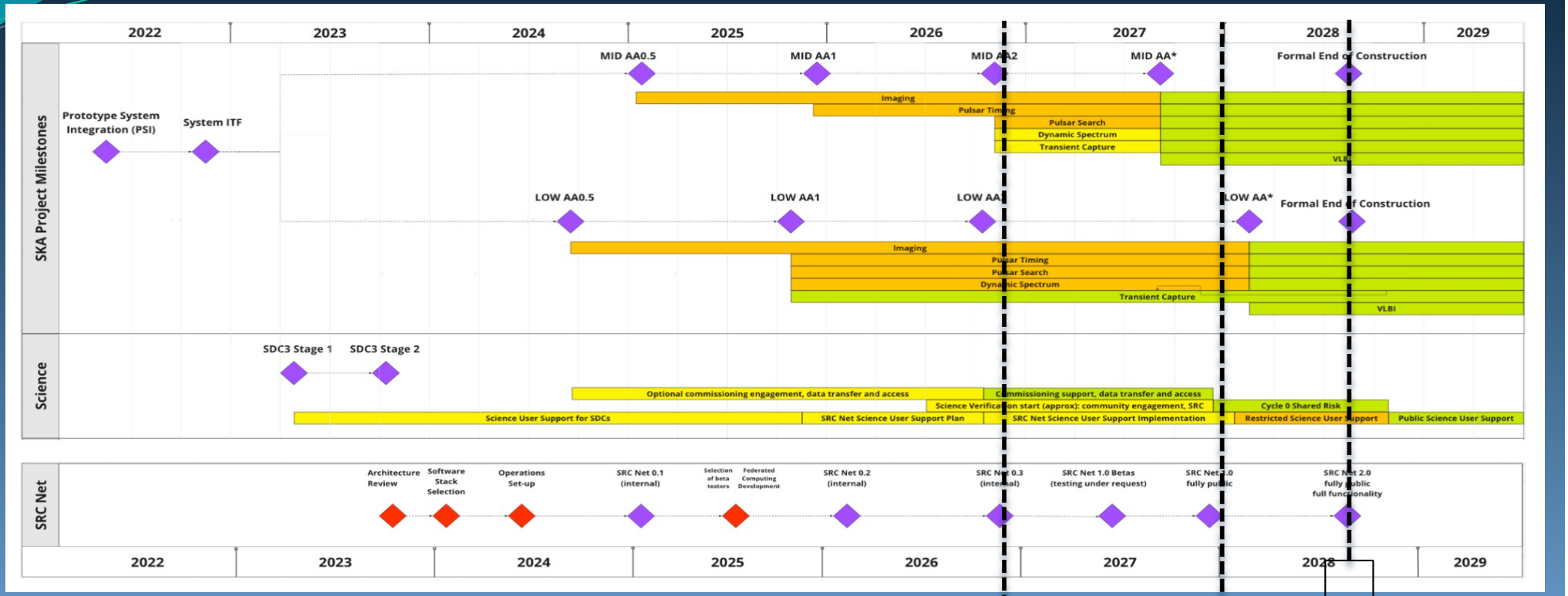
- 1.a Data products replication, distribution, and synchronization across multiple locations (Rucio)
- 1.b Data products replication, distribution, and synchronization across multiple locations (CADC)
- 2. Federated Authentication and Authorization (AA) API
- 3. Data processing Notebooks
- 4. Visualization of SKA data with high volume of users and high amount of data
- 5. Distribution of software, tools and services

The Epics

EPIC name	Content & scope	Inaf involvement
System Sizing for the SRCNet	Understand hardware requirements from current pre-SKA data processing/archiving and to scale those requirements to the actual SKA data processing/archiving	
SRCNet AAI Design	Create design for SRCNet AAI	
Software distribution system	Set up a system to provide software to nodes.	
SDC3b science platform testing	Aim to have packaged science platform ready and deployed at multiple sites to show (1) ease of deployment, (2) integration across multiple services (3) gather user feedback from: i) operators, ii) science team as SDC3b administrators, and ideally iii) science data challenge participants	
Mini-SRCNet demonstrator	An end-to-end multi-SRC node demonstrator from data management to execution planning with integrated A&A and interoperability across all sites	Fabio Vitello
Federated services	Understand how we can start to deploy services across different physical infrastructure	
Development and collation of runnable example workflows	Develop a plan for data movement tests and challenges	
Data Lake integration	Improve Rucio integration with data discovery tools with IVOA metadata integration, compute platforms, and visualization tools.	Fabio Vitello
Computing Services API	Develop a client-server API to submit/execute operations on the SRC federated computing resources and for the related monitoring. This EPIC will also cover the retrieval of information	**Claudio Gheller** (lead) Giuliano Taffoni Sara Bertocco Fabio Vitello

it SRC
SKAO Regional Centre Italy

The Roadmap of the SRCnet



Preliminary Plan

SRCnet available to support AA2

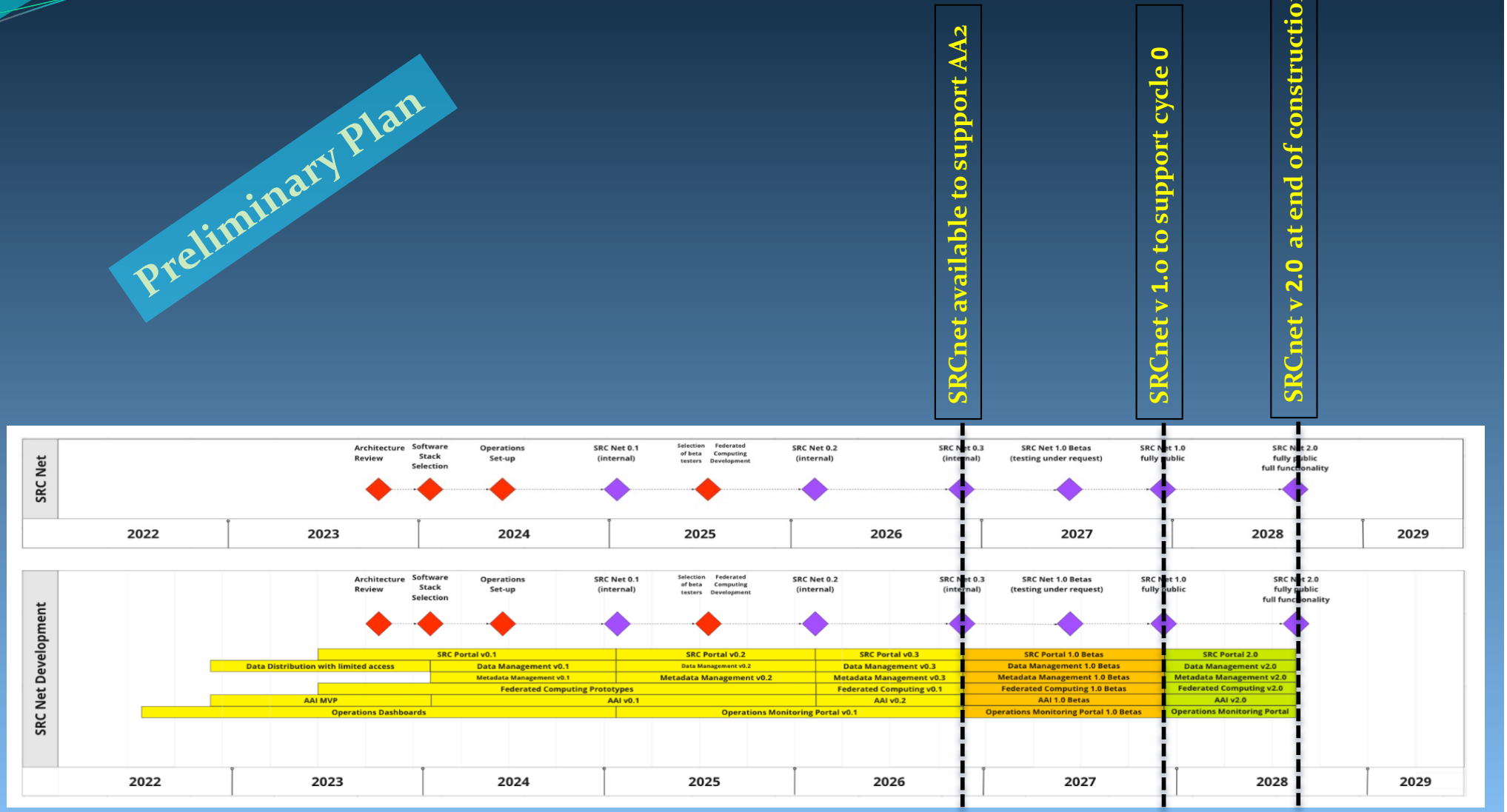
SRCnet v1.0 to support Cycle 0

SRCnet v2.0 at end of construction

it | SRC
SKAO Regional Centre Italy

The Roadmap of the SRCnet

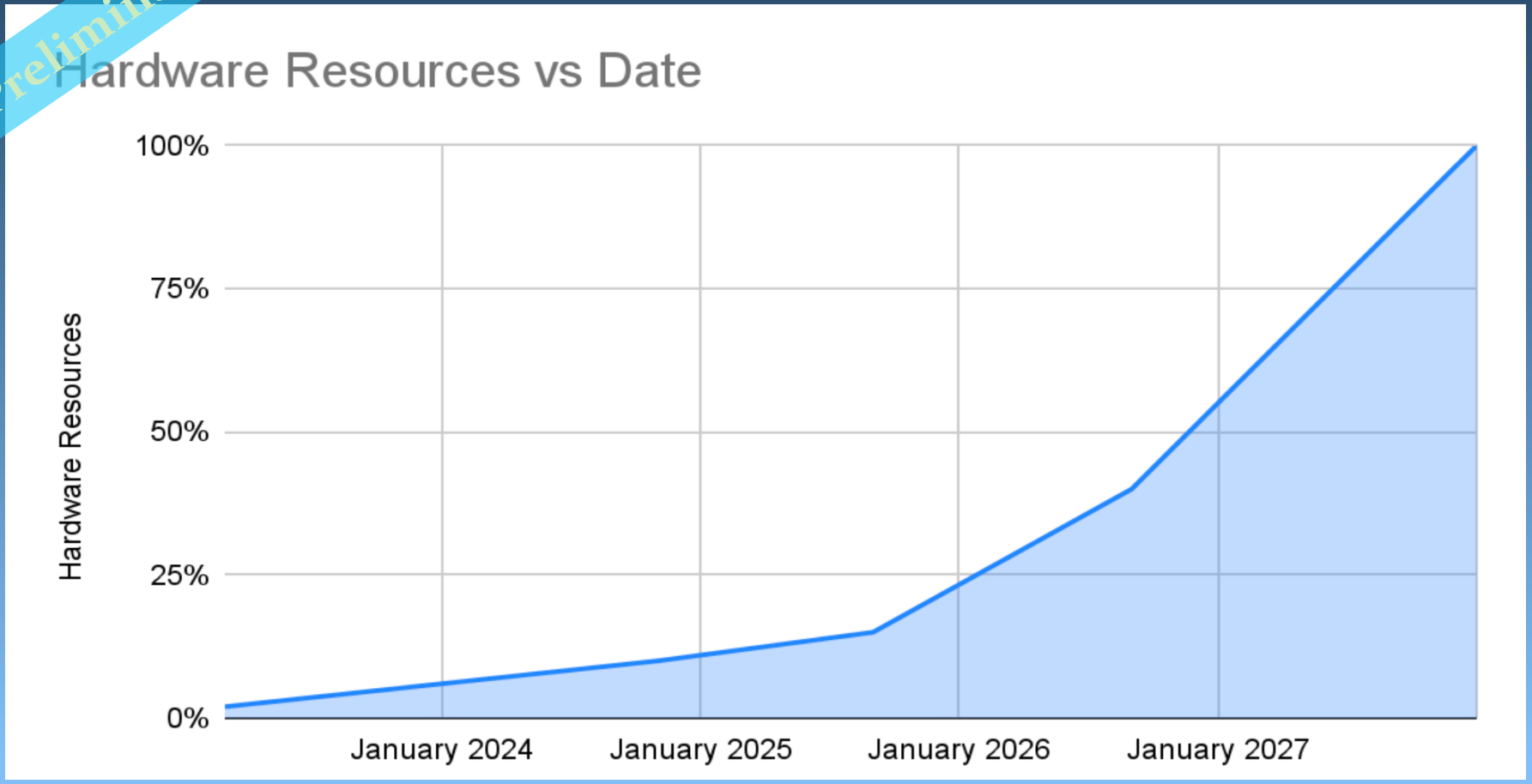
Preliminary Plan



it SRC
SKAO Regional Centre Italy

Hardware resources growth

Preliminary Plan



it SRC
SKAO Regional Centre Italy

Some early estimates

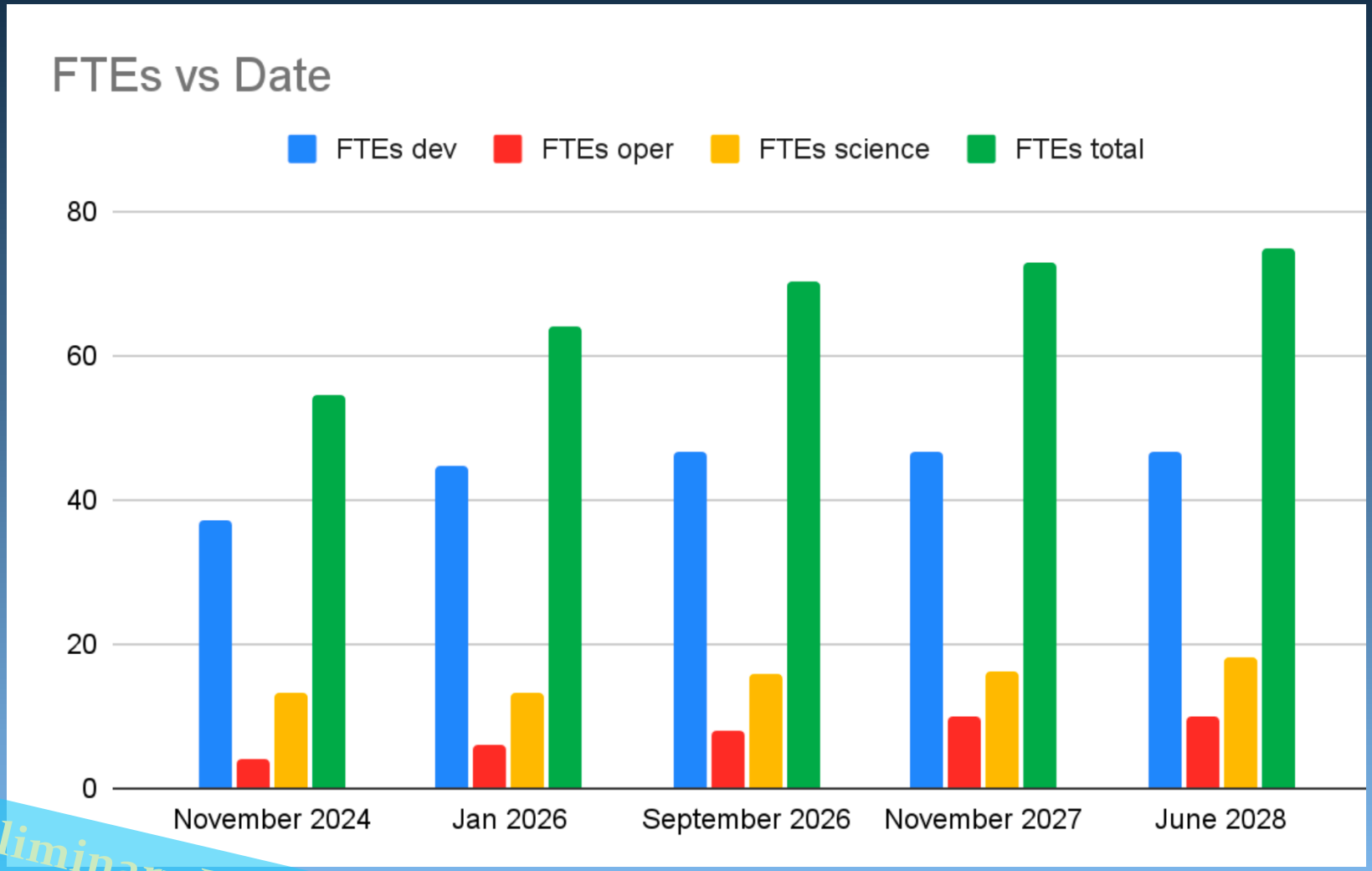
Preliminary Plan

		SRC Net v0.1	SRC Net v0.2	SRC Net v0.3	SRC Net v1.0b	SRC Net v1.0
		Jan 2025	January 2026	Sep 2026	Nov 2027	Jun 2028
Deployment (%)		2.00	10.00	15.00	50.00	100.00
Country	Share (%)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)
Italy	6	**0.04**	**0.21**	**0.32**	**1.05**	**2.10**
Total	100	0.70	3.50	5.25	17.50	35.00

		SRC Net v0.1	SRC Net v0.2	SRC Net v0.3	SRC Net v1.0b	SRC Net v1.0
		Jan 2025	January 2026	Sep 2026	Nov 2027	Jun 2028
Deployment (%)		2.00	10.00	15.00	50.00	100.00
Country	Share (%)	Storage (PB)	Storage (PB)	Storage (PB)	Storage (PB)	Storage (PB)
Italy	6	**1.27**	**6.36**	**9.54**	**31.80**	**63.60**
Total	100	21.20	106.00	159.00	530.00	1060.00

it | SRC
SKAO Regional Centre Italy

Human resources growth



Preliminary Plan

it SRC
SKAO Regional Centre Italy

Italian expected outcome of the SRC network foundation

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

✓ 2. The establishment of a SRC network with a significant node located in Italy

✓ 3. The recognition of the local investments in both hardware and human expertise, and its conversion into incentives as soon as possible

it | SRC
SKAO Regional Centre Italy

The Italian contribution to the global effort

science

≈ 100 Italian astro-scientists are members of the SKA Science Working Groups!

Develop requests and imagine solutions to the USE CASES for the SRC network

+

Staying at the frontline in ADAPTING to the new way for doing data reduction and computation in the SKA era

+

Exploit experience in precursors & pathfinders to provide suggestions and solutions

SKA Science Regional Centres - SCSRC community input

Survey Flow

Standard: Questionnaire Preamble (2 Questions)
Standard: Section 0 - Some general questions (6 Questions)
Standard: Section 1. Data products and scientific requirements (17 Questions)
Standard: Section 1. Data products and scientific requirements_Loop (66 Questions)
Standard: Section 2. Archive mining and VO Interface (19 Questions)
Standard: Section 3. Post-processing – Analysis – Visualisation (53 Questions)
Standard: Section 4. User support (11 Questions)

The Questionnaire for the SWGs: 174 questions!

it | SRC
SKAO Regional Centre Italy

The Italian contribution to the global effort

expertise

Role	First Name	Family Name	Agile Teams
Business Owner	Claudio	Gheller	all
Product Owner	Fabio	Vitello	Orange
Scrum Master	Giuseppe	Tudisco	Orange
member	Eva	Sciacca	Orange
member	Andrea	Lorenzani	Orange
member	Alessandra	Zanichelli	Orange
member	Vincenzo	Galluzzi	Orange
member	Robert	Butora	Orange
member	Marco	Molinaro	Orange
member	Gianluca	Marotta	Orange
member	Gianmarco	Maggio	Olive
member	Sara	Bertocco	Olive
member	Matteo	Di Carlo	Olive
member	Matteo	Canzari	Tangerine
stakeholder	Matteo	Stagni	Cyan (not active)
stakeholder	Cristina	Knapic	Cyan (not active)
member	Giuliano	Taffoni	Architecture Group

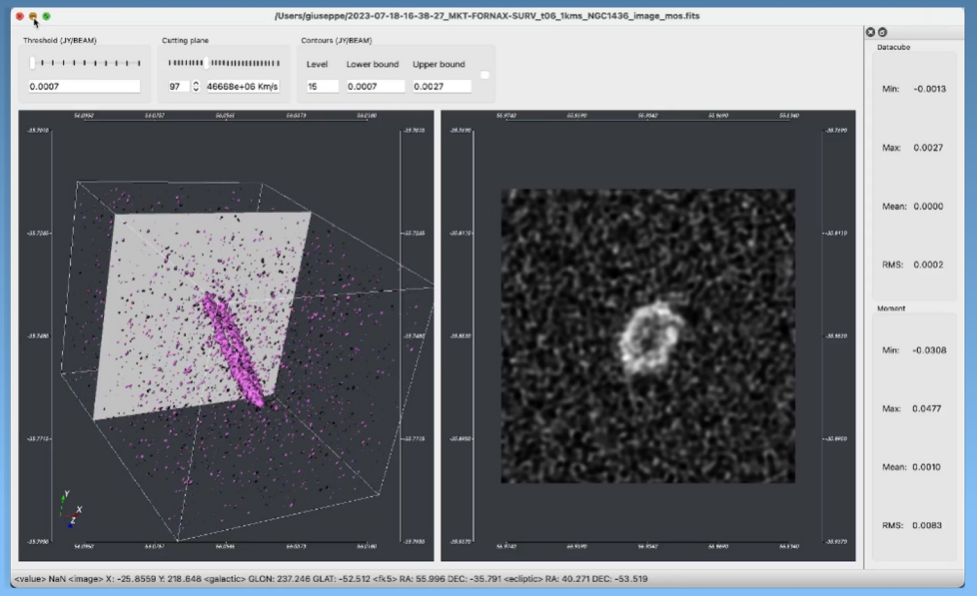
Country	Total anticipated fractional FTE for PI21 (SRCNet global activities)
Australia	0,00
Canada	2,00
Switzerland	1,00
China	4,60
France	0,80
Germany	0,00
India	0,00
ITALY	**2,15**
Japan	1,20
Korea	0,80
Neaderland	3,05
Portugal	0,00
South Africa	0,00
Spain	4,10
Sweden	0,80
United Kingdom	4,35
SKAO	5,30
Total	30,15

it | SRC
SKAO Regional Centre Italy

The work of the Orange Team

Started with working on Prototype 4: Visualization in PI15 (June 2022)

- Contributing to the **definition** of visualization **use cases** for SRCNet
- **Visualization Tools review** (dependencies, interfaces, workshop)
- **Collection of data products** and data formats from **precursors and pathfinders**
- **Adapting Visualization Tools to address use cases** and work with SRC architecture and its data lake
- Development, testing and deployment of **SODA** (Server-side Operations for Data Access) into SRCNet, integrated with **Rucio Data Lake** and **Discovery services**
- Review of Solutions and Technologies for the **Computing Services API**
- Testing and **deployment** of visualization tools and data access services **into SRC nodes**



Courtesy Tudisco & Vitello

it SRC
SKAO Regional Centre Italy

The work of the Orange Team

the next steps

Contributions on the following SRCNet **Epics**

- **Mini SRCNet Demonstrator**
  - Benchmark and Optimization of SODA performances
- **Data Lake Integration**
  - Adapt SODA to work with storage manager service (SRM)
  - Collection and classification of heterogeneous datasets from precursors and pathfinders for testing services and tools
- **Computing API**
  - Adapt SODA so that it can process requests through the Computing API (Application Programming Interface)
  - Adapt VisIVO to invoke cutout and visualize data through the Computing API (Application Programming Interface)

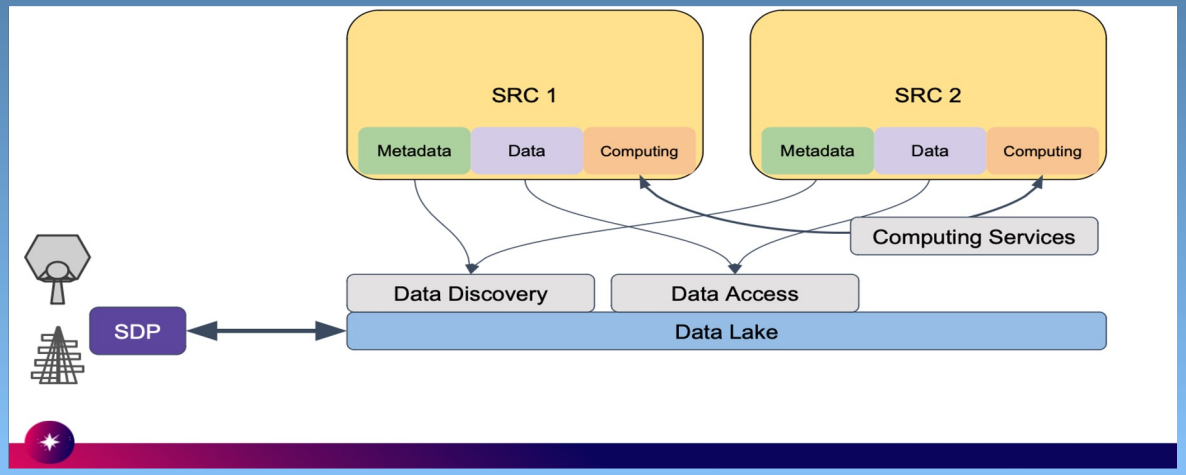


image credits:Jesús Salgado - SKA Regional Centre Architect

Courtesy Tudisco & Vitello

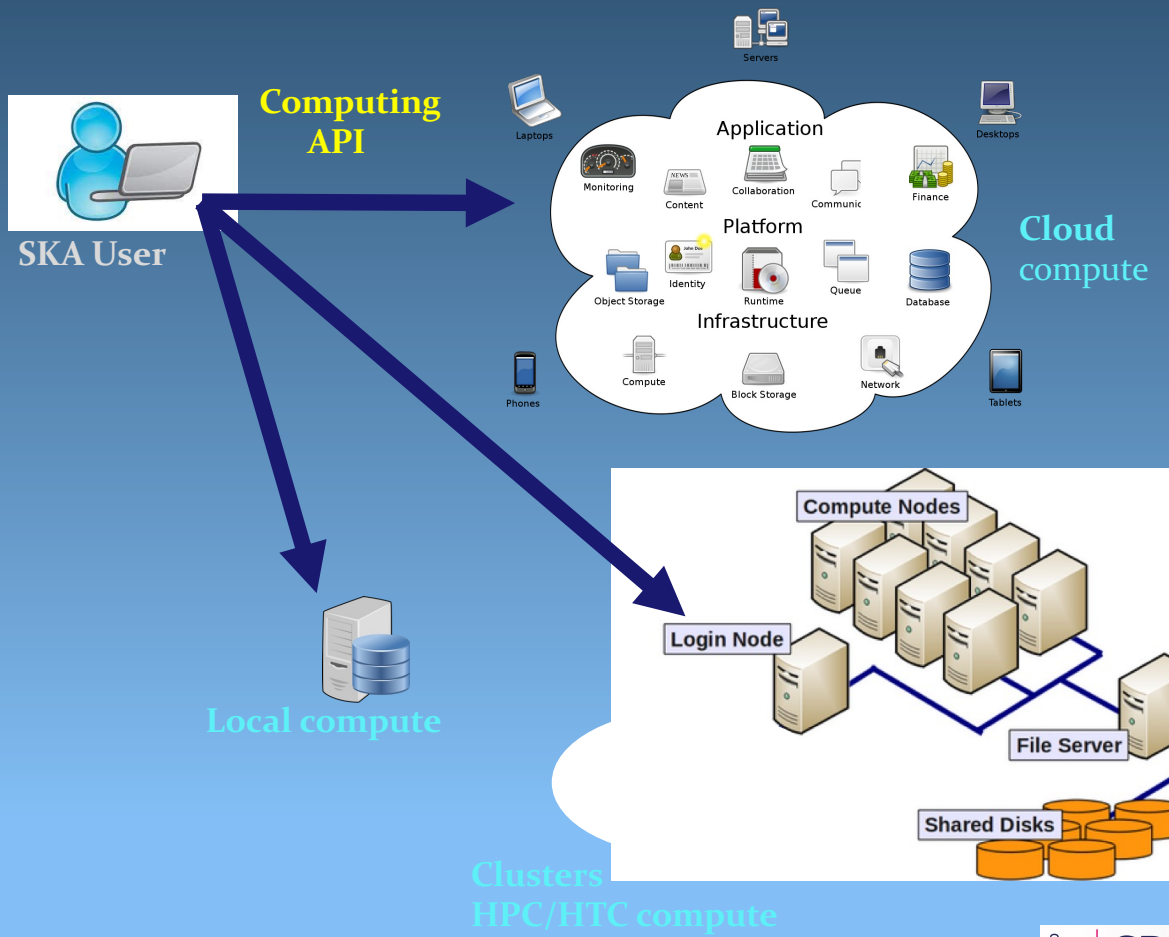
it SRC
SKAO Regional Centre Italy

The work of the Olive Team

Contributions on Computing API (Application Programming Interface)

Discussion

- Use cases
- Requirements
- Authentication and Authorization
- Solutions
  - ExecutionPlanner ?
  - Dirac ?
  - Others ?



Courtesy Bertocco & Maggio

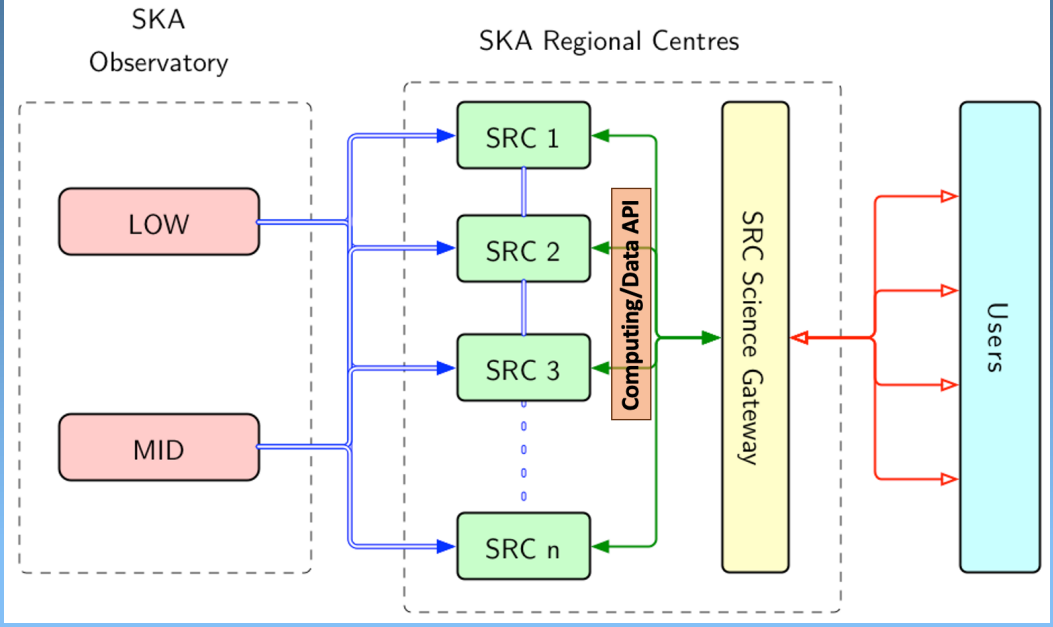
it SRC
SKAO Regional Centre Italy

The Epic about Computer Service API

(Application Programming Interface)

Aim: design & implement an Application Programming Interface (API) to

1. Discover computing services, answering the question "what computing services are available and suitable to run my task?"
2. Access computing services, answering the question "how can I run my task on the selected computing service (and when)?"



THE FOCUS IS NOT ON THE TECHNOLOGY, BUT TO PROVIDE AN ABSTRACTION ABOVE TECHNOLOGY

Courtesy Gheller

it SRC
SKAO Regional Centre Italy

The Italian contribution to the global effort

hardware: initial assets

Acquisition of **≈1.5 PetaFlop/s** (with a combination CPU and GPU) **and ≈11 PBy** (combined between fast disks for computing and tapes for long-term preservation) **Tier-3** computing system, to be installed inside of one of the CINECA areas at **the Bologna Technopole**

The Technopole already hosts the European weather centre ECMWF, the Leonardo super-computer and will host the United Nations University on Climate Change

Use of **a Tier-2 sizing system** integrated into a Tier-1 sizing system and becoming the **kernel of the Italian node of the SKA Regional Center.** Investment of the CN-PNRR for the needs of INAF and CNR owned by CINECA, with guaranteed (non-exclusive) use for INAF. Expected for INAF: **about 4 PetaFlop/s** (Data Centric Nodes and Booster Nodes) and **≈2 PBy** high speed storage

it | SRC
SKAO Regional Centre Italy

The development of the INAF node

hardware on the longer term (beyond 2024)

1st **Step in the Global effort: We will begin** with a **Tier-3/Tier-2** protoSRC **by the end of the 2024** (**v 0.1** of the SRCnet node)

2nd **Step in the Global effort: … and progressively attain, by mid 2028**, a **Tier-2/Tier-1.5** size system (**v 1.0** of the SRCnet node) with capability of ≈ **2 Pflops** and ≈ **60 PBy** of storage (30 PBy on-line and 30 PBy near on-line), connected at **100 GB/s** with the other nodes

3rd **Step in the Global effort: … and finally, by 2030**, a **Tier-1** size infrastructure with capability of ≈ **3+ Pflops** and ≈ **150 PBy/yr** of storage (45 Pby on-line and 105 Pby/yr near on-line), connected at **100 GB/s** with the other nodes (**v 2.0** of the SRCnet node)

Technopolo - Bologna

National effort: growing in parallel an infrastructure with capability of ≈ **5+Pflops** and ≈ **10 PBy/yr** of storage connected at **100 GB/s** with the other nodes for the national needs related to precursors/pathfinders and to exploitSKA data and SKA archives

it | SRC
SKAO Regional Centre Italy

Roadmap of the INAF node: the national effort

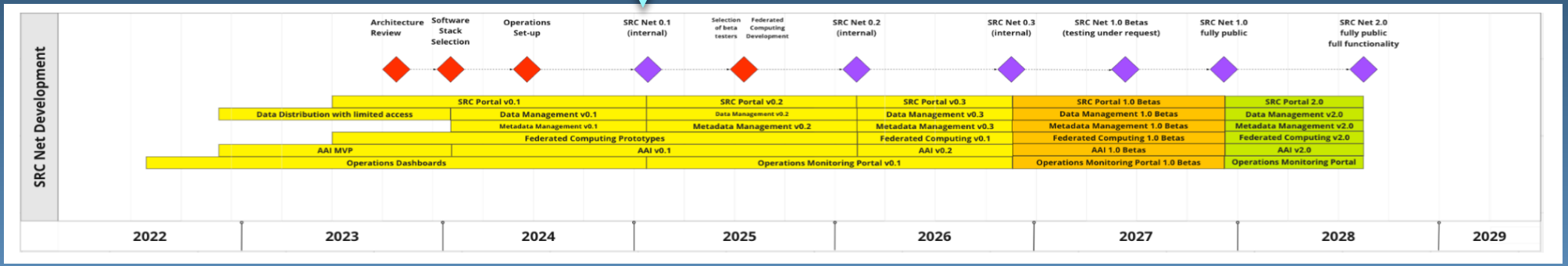
First Name	Family Name
Claudio	Gheller
Fabio	Vitello
Giuseppe	Tudisco
Francesco	Bedosti
Alessandra	Zanichelli
Vincenzo	Galluzzi
Marco	Molinaro
Gianmarco	Maggio
Sara	Bertocco
Matteo	Stagni
Cristina	Knapic
Giuliano	Taffoni

it | SRC
SKAO Regional Centre Italy

Roadmap of the INAF node: the national effort

SRC.IT prototypization

Objective: create a first prototype of Italian SKA Regional Center integrated in SRCnet v 0.1



On-going tasks:

- Set-up of a data-lake based on object storage solutions (S3+CEPH)
  1. Deploy RUCIO on the top of the distributed repository
  2. Connect to the RUCIO based SRC data infrastructure
  3. Experiment VLKB solution for metadata management

- Deployment of computing services for on site data-local analysis and visualization
  1. Access through the Computing API
  2. Analysis af SKA supported AAI approaches
  3. Remote visualization

Courtesy Gheller

it SRC
SKAO Regional Centre Italy

Global and national effort for the INAF node

the need of personnel

- to match the Italian percentage involvement in the SKAO construction, at least 6 FTE must be devoted to the SRC network at regime
- to run a significant **Italian node** of the SRC network, **about 10 FTE** are needed at regime

The role of this team will mainly be

(a) **understanding** of the operations of the data acquisition systems,
(b) **operating,** jointly with SKAO, the network that will oversee the data analysis/curation/archiving,
(c) **developing** of scientific software, with the fundamental aid of the national and international scientific community
(d) **interacting with** and help the users in the preparation and management of the observing programs

it | SRC
SKAO Regional Centre Italy


Thank you !
IN PARTNERSHIP WITH SKAO
INAF
ISTITUTO NAZIONALE DI ASTROFISICA
USC VIII
it | SRC
SKAO Regional Centre Italy
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Finanziato dall'Unione europea
NextGenerationEU
Thank you !