

I requisiti degli *SKA Regional Centers* per una piattaforma federata su Cloud

Eva Sciacca

INAF, Osservatorio Astrofisico di Catania

WORK
SHOP
GARR
2021

NET
MAKERS



INAF

ISTITUTO NAZIONALE
DI ASTROFISICA

Sommario

Il progetto SKA

SKA Regional Centers

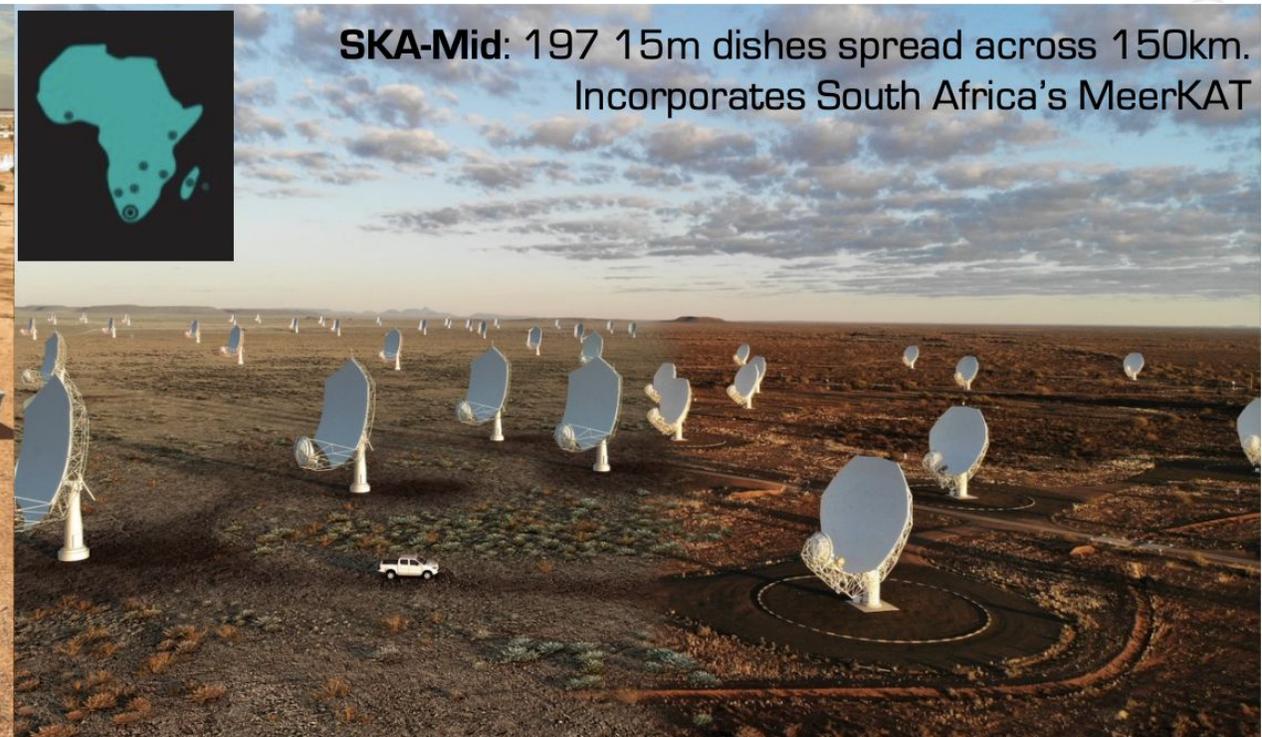
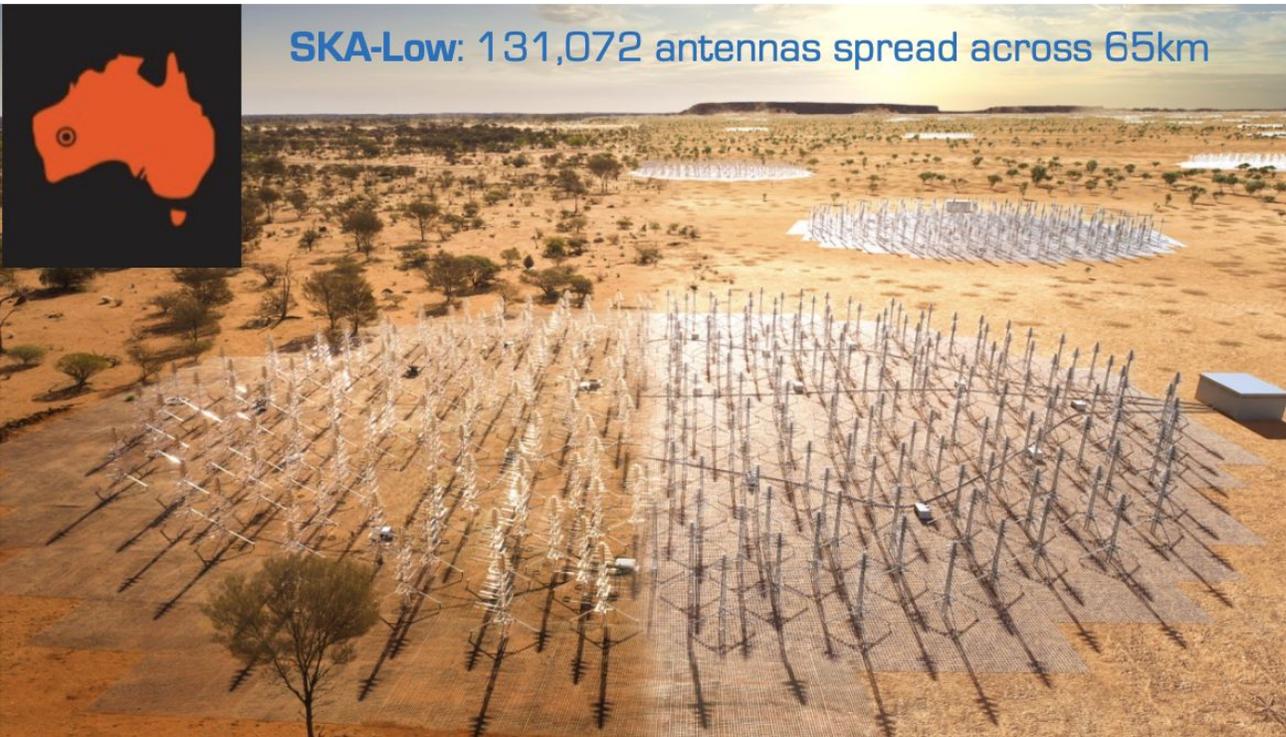
Requisiti Piattaforma Federata

Prototipazioni @ Cloud GARR

- Servizi di federazione
- Visualizzazione
- Analisi

Il Progetto SKA

- Una collaborazione di 16 paesi sta costruendo l'**osservatorio radioastronomico** più grande al mondo.
- Gestito dallo **SKA Observatory** (SKAO), organizzazione intergovernativa dal 2021, con sede presso Jodrell Bank (a Manchester) nel Regno Unito.



Flusso dati di SKA

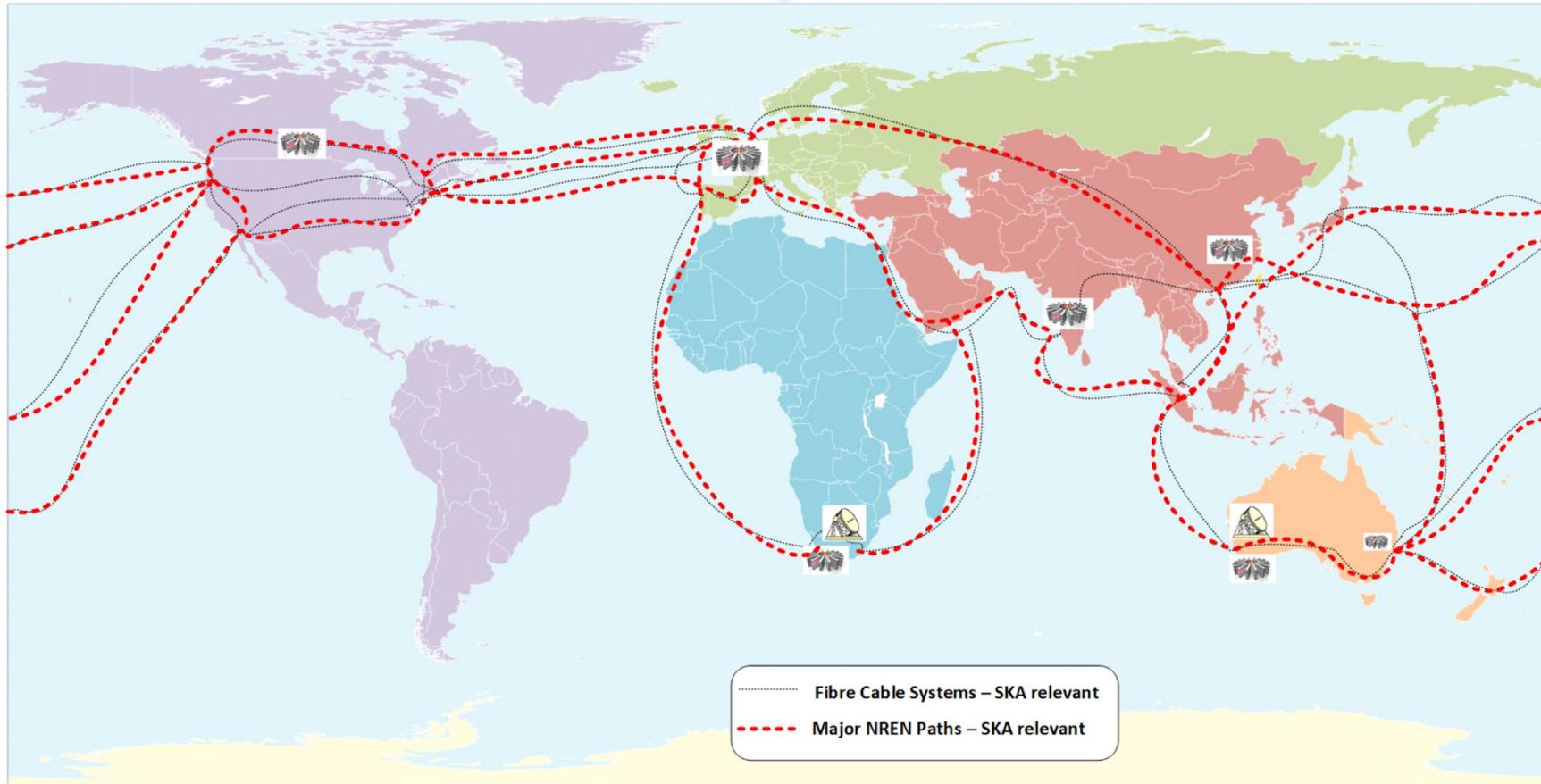


La velocità dei dati prodotti da SKA sarà approssimativamente più alta di 50-70 volte di altri telescopi che osserveranno la volta celeste in dettagli mai raggiunti fino ad ora, come il l'Osservatorio Vera Rubin (LSST).

I dati prodotti da SKA non saranno nello stato finale richiesto per l'analisi scientifica e il loro volume sarà così grande che la consegna diretta agli utenti finali, distribuiti geograficamente su scala mondiale, è irrealizzabile.



SKA Regional Centers - SRC



- Pianificazione di **una rete di SRC** in tutto il mondo che ospiteranno lo *SKA Science Archive*
- Gli utenti potranno accedere lo SKA Science Archive tramite il relativo SRC
- Gli SRC forniranno **risorse** per ulteriori **elaborazioni** e **analisi** dei dati
- Gruppi di lavoro stanno sviluppano la **progettazione** degli SRC



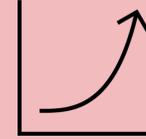
Per maggiori info su L'OSSERVATORIO SKA E GLI SKA REGIONAL CENTRES si riveda l'intervento di Andrea Possenti (INAF) di Lunedì 8 Novembre.

Piattaforma Federata – Design Principles



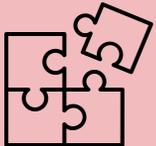
Generico

- Il sistema deve essere in grado di supportare **più team scientifici** utilizzando i dati dei telescopi precursori di SKA e, infine, supportare la scienza di SKA.



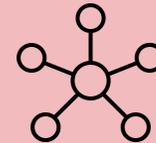
Scalabile

- Il sistema deve tenere conto dei **crescenti requisiti di elaborazione dei dati** da parte della comunità astronomica.



Estensibile

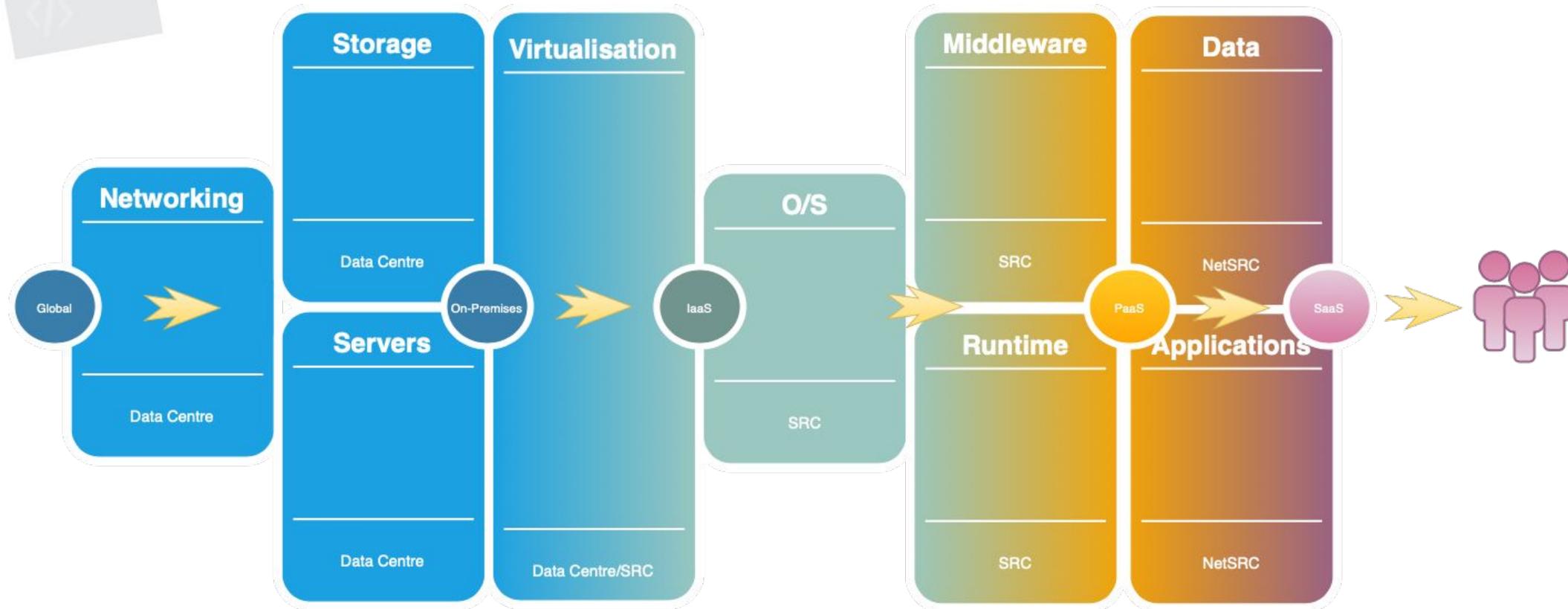
- Dovrebbe essere possibile **aggiungere servizi** in un secondo momento.
- Dovrebbe essere **distribuito** in tutto o in parte su diversi fornitori di cloud e data center.



Debolmente collegato

- I servizi nel sistema dovrebbero essere **sufficientemente indipendenti**.

Piattaforma Federata - SaaS



- L'obiettivo è quello di abilitare e semplificare il processo di fare scienza con grandi set di dati di radioastronomia attraverso il concetto di "portare il processamento ai dati" o In-Archive-Processing (IAP).
- Concetto di sistema basato su cloud Software as a Service (SaaS) fornisce **servizi persistenti comuni** e **specifici per la scienza**, nonché archiviazione, elaborazione, database e altri servizi disponibili su richiesta a un utente finale per eseguire la post-elaborazione e l'analisi dei dati.

Piattaforma Federata – Requisiti (SRC WG3)

- Sarà disponibile una **soluzione federata per l'accesso** (Infrastruttura di autenticazione ed autorizzazione AAI) che abilita il comportamento SSO (Single Sign-on).
- L'SRC fornirà servizi per **visualizzare** (e **analizzare**) immagini con dimensioni di file fino a 1 PB.
- L'SRC fornirà agli astronomi una piattaforma per progettare, **prototipare**, implementare e testare il software SRC e i workflow in modo **interattivo**.
- L'SRC fornirà funzionalità di **aggregazione** attraverso la quale i servizi federati possono registrare centralmente informazioni accumulate (ad esempio, tenere traccia dell'utilizzo delle risorse e calcolare il tempo per utente o per gruppi di utenti).
- L'SRC fornirà un sistema di **logging** attraverso il quale i servizi SRC federati possono accumulare log, generati in modo distribuito, in unico archivio centralizzato.
- ...

Piattaforma Federata - Architettura di alto livello

Visualization and Analysis Portal

SRCNet API

SRCNet Federated Services

Authentication and Authorization

Data Mgt

Computing Mgt

Accounting/
Logging etc.

Prototipazioni @ GARR Cloud



Visual Analytic



Visualization Gateway



Source Finding and Deep/Machine Learning

API

API

API

Core Services



Authentication and Authorization



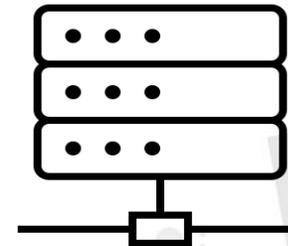
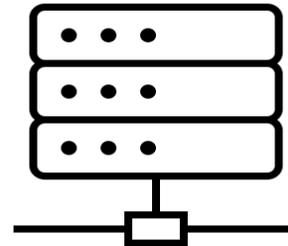
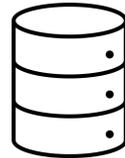
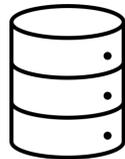
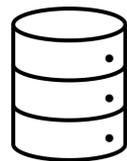
Data Mgt



Computing Mgt



Accounting/Logging etc.





AAI

- Offre una **soluzione orizzontale** per l'autenticazione a tutti i servizi
- Supporta un certo grado di **autorizzazione** di queste richieste
- Basato su **protocolli e standard** ampiamente accettati per garantire un'ampia applicabilità dell'approccio.



Sign in with

 NEANIAS SSO

Remember me



NEANIAS

English v

Username or email

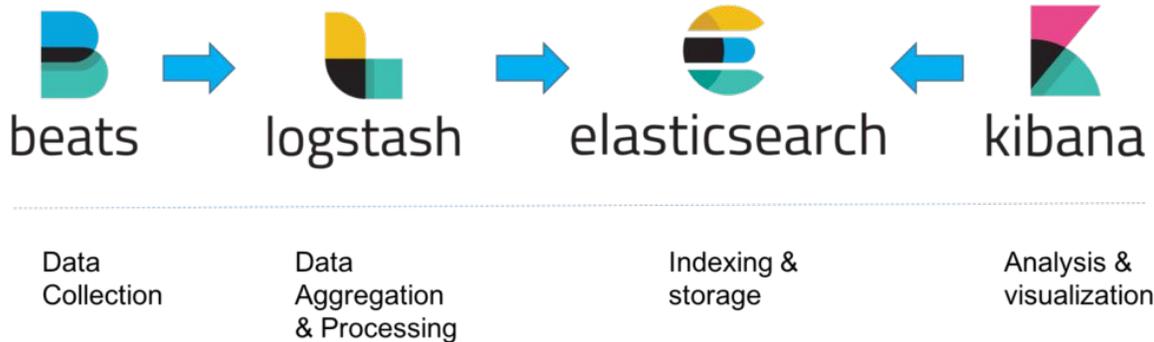
Password

Log In

Login with NEANIAS

Logging & Accounting

- Il servizio di Logging fornisce funzionalità di aggregazione.
 - I servizi federati accumulano log generati in modo distribuito in un unico repository centralizzato.
- Il servizio di Accounting fornisce funzionalità di contabilità.
 - I servizi federati registrano centralmente le informazioni aggregate man mano che vengono accumulate gradualmente.



Dashboard: Server status

DLQ Status: Count of records vs entry_time per day

All Services Status: Count of records vs timestamp per day

All services logs: Count of records vs timestamp per day

Accounting Invalid Data: Count of records vs @timestamp per day

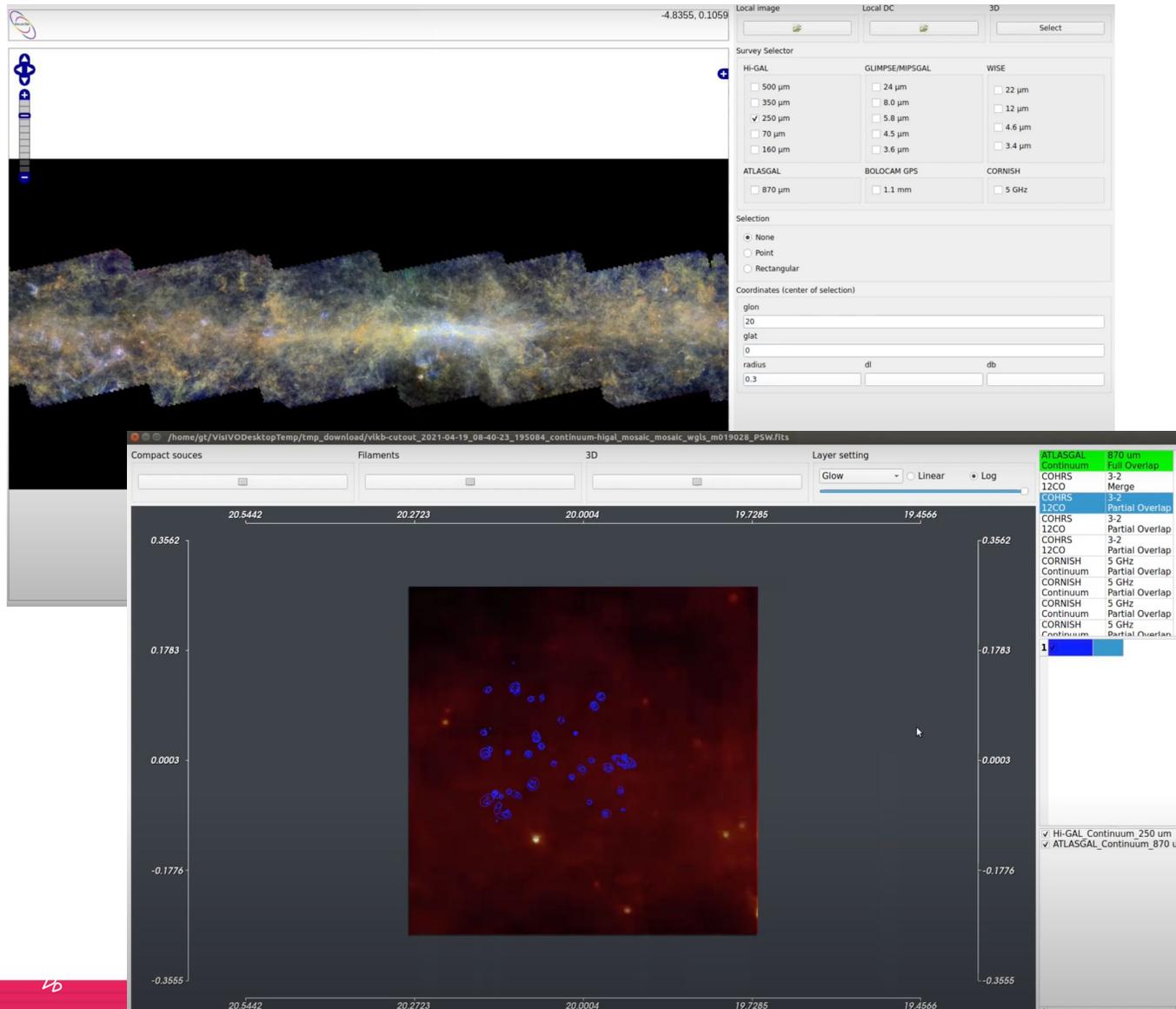
Accounting Entries Per Service: Count of records vs timestamp per day

Users Table:

Subject	Name	Email	Service	Parent	Created
41ee1bba-8539-46d9-97da-dcfb3d54f0e7 (Robert Butora)	41ee1bba-8539-46d9-97da-dcfb3d54f0e7 (Robert Butora)		vikb		5/15/21, 9:59 AM
43dee190-80b4-4729-8d6d-fb6b592e5ba (Fabio Roberto Vitello)	43dee190-80b4-4729-8d6d-fb6b592e5ba (Fabio Roberto Vitello)		vikb		6/8/21, 2:04 PM
44c29ebf-0287-4773-a7eb-389c60594a0f (Silvia Tofani)	44c29ebf-0287-4773-a7eb-389c60594a0f (Silvia Tofani)		vikb		6/16/21, 10:02 AM



ViaLactea Visual Analytics - VLVA



- Fornisce l'accesso ai **survey radio e infrarossi** della **nostra galassia** archiviati nella Knowledge Base (VLKB)
- Supporta la visualizzazione di immagini 2D e dati di velocità 3D (vol. rendering, sezioni), caricati localmente o da VLKB
- Abilitazione della visualizzazione di sorgenti compatte e filamentari dal VLKB
- Disponibile come **applicazione desktop** per Mac OS e Ubuntu e anche come **container Docker**.

Maggiori dettagli @ Vitello, F., et al.

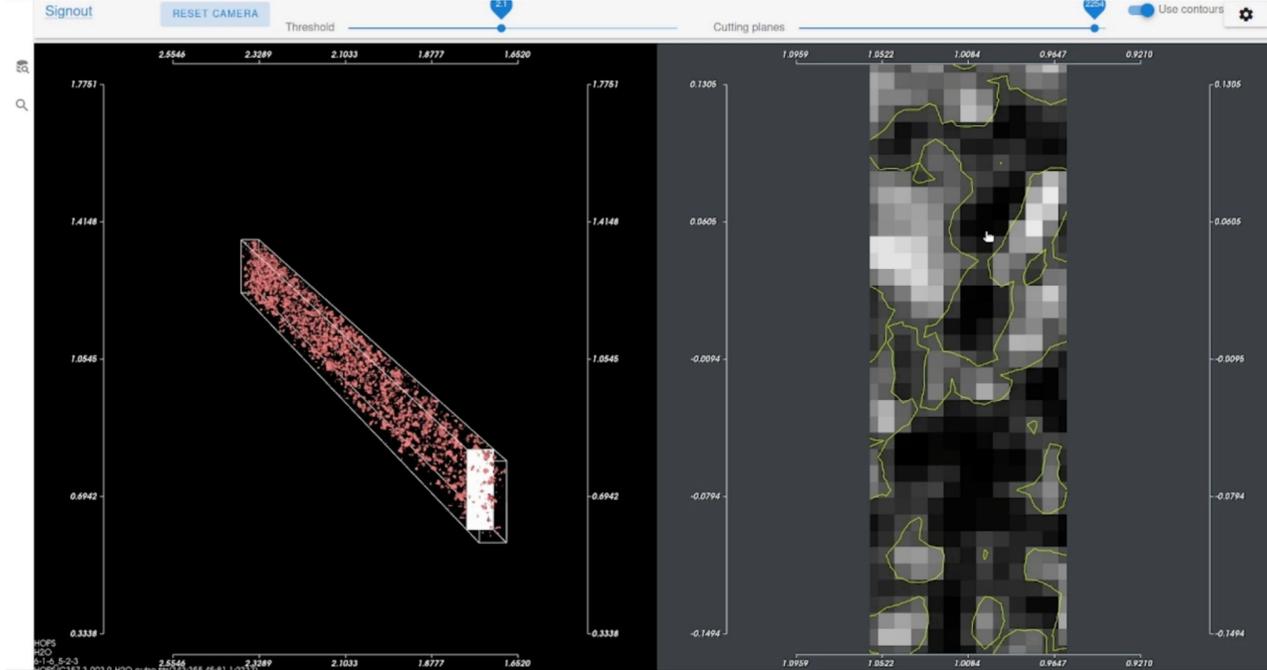
Publications of the Astronomical Society of the Pacific (2018)

<https://github.com/NEANIAS-Space/ViaLacteaVisualAnalytics>





ViaLactea Web - VLW



- Versione **web semplificata** del VLVA, sviluppata in collaborazione con l'Università di Portsmouth (UK)
- Attualmente supporta solo la visualizzazione di cubi di dati di velocità 3D
- Supporto multiutente all'interno dell'ambiente web
- Fornisce un' **efficiente visualizzazione** (rendering GPU e CPU) su server remoto

- Soluzione browser Web con supporto desktop/mobile/tablet flessibile per future estensioni funzionali
- Stiamo lavorando sulla **containerizzazione completa** per una **distribuzione ottimizzata su cloud** e infrastrutture distribuite



ViaLactea Knowledge Base - VLKB

The screenshot displays the 'Table Access Protocol (TAP) Query' interface. It features a 'Metadata' tree on the left with a search bar and checkboxes for 'Name' and 'Descrip'. The right panel shows details for the selected table 'compactsources.sed_view_final', including its name, columns (203), foreign keys (0), and description: 'VIALACTEA Catalogue of band-merged source fluxes'. Below this, there are sections for 'Service Capabilities' (Query Language: ADQL-2.0, Max Rows, Uploads) and 'ADQL Text' (Mode: Synchronous). The ADQL query editor contains the text: 'SELECT TOP 1000 * FROM compactsources.sed_view_final'. At the bottom, there are 'Examples' and 'Basic 1/6: Full table' tabs, and a 'Run Query' button.

- Fornisce servizi di ricerca ed accesso a raccolte di dati e cataloghi del piano galattico.
- Accessibile attraverso un'infrastruttura abilitata al Virtual Observatory (tramite protocollo TAP).
- Accesso ai dati disponibile tramite servizi API REST: ricerca, ritaglio e unione di mappe.
- Fornisce inoltre informazioni su sorgenti compatte, strutture di filamenti e modelli numerici
- Completamente sfruttato dagli strumenti client di ViaLactea Visual Analytic (desktop e web).
- Ospitato presso GARR Open Stack
- Protetto con l'infrastruttura di autenticazione e autorizzazione (AAI)
- Gestisce i ruoli di privacy degli utenti relativi all'accesso a specifiche survey





Visualization Gateway - VG

- Basato sul framework **JupyterHub** e include **VisIVO** e **Splotch** come ambienti di visualizzazione.
- Servizio di **data sharing** integrato con montaggio automatico contenente dati demo (sola lettura).
- In esecuzione sul cluster GARR Kubernetes

Maggiori dettagli @ E. Sciacca et al (submitted) *Journal of Grid Computing*

Server Options

Remote storage mode

Auto-mount
Auto-mount to remote reference-data filestore (Nextcloud)

Other
Mount to other filestore via WebDAV

VD-Splotch
VD-Splotch command line tools

VD-VisIVO
VD-VisIVO command line tools

jupyterhub demo-colours (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Code

```
In [4]: # image3d
path="image3d/image3d.fits"
d="image3d"

!VisIVOImporter --format fitsimage --vol
!VisIVOViewer --volume --vrenderingfield $datasetPath/$path --showbox --showaxes --showlut VisIVOserverBinary.bin
mkdir $d
mv VisIVO* $d

In [5]: # When the cell above has finished executing, run this cell to see the output image.
from IPython.display import Image

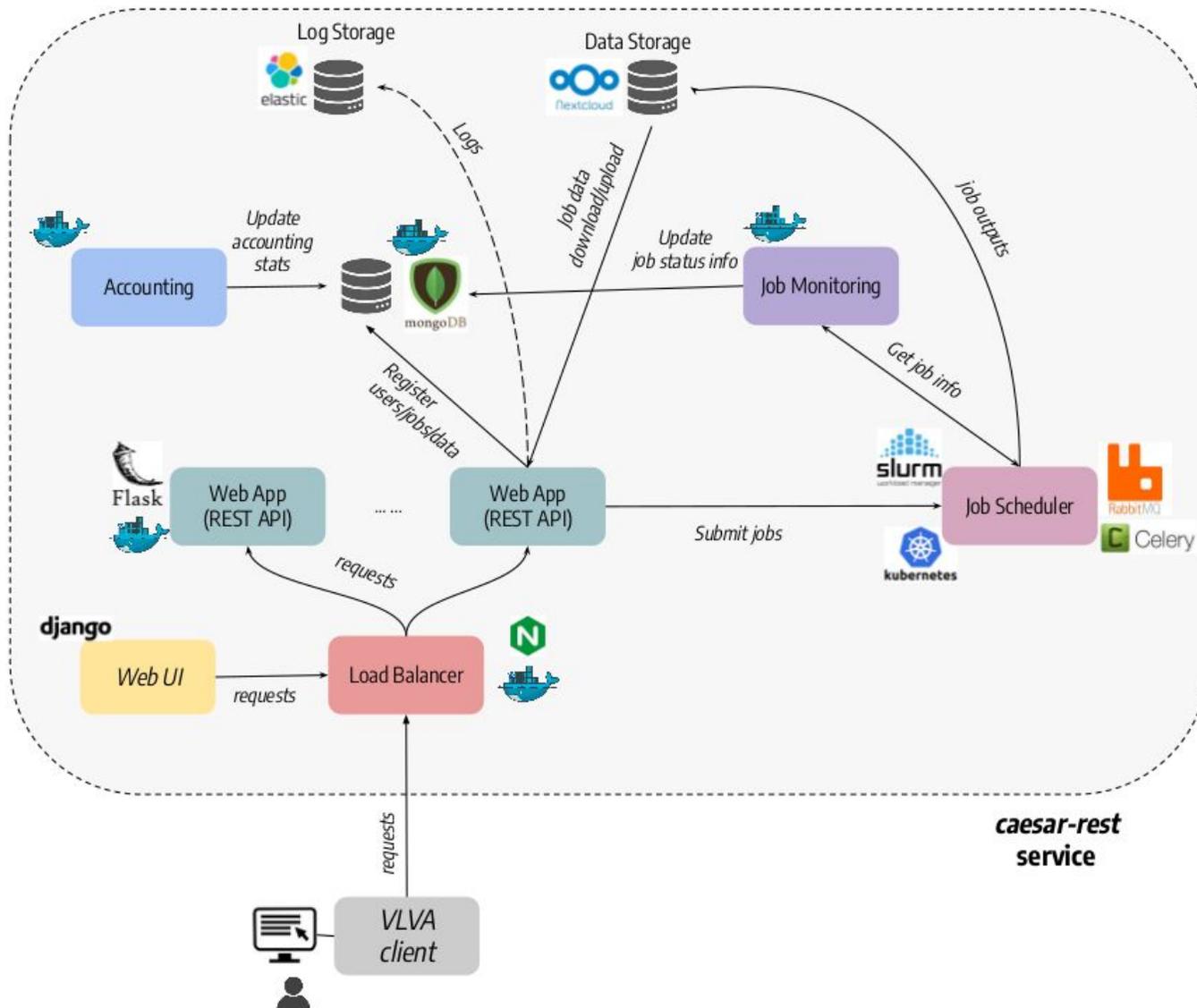
# Display the image produced
imageFile = d + "/VisIVOserverImage3.png"
Image(filename=imageFile)
```

Out[5]:





SOURCE FINDING



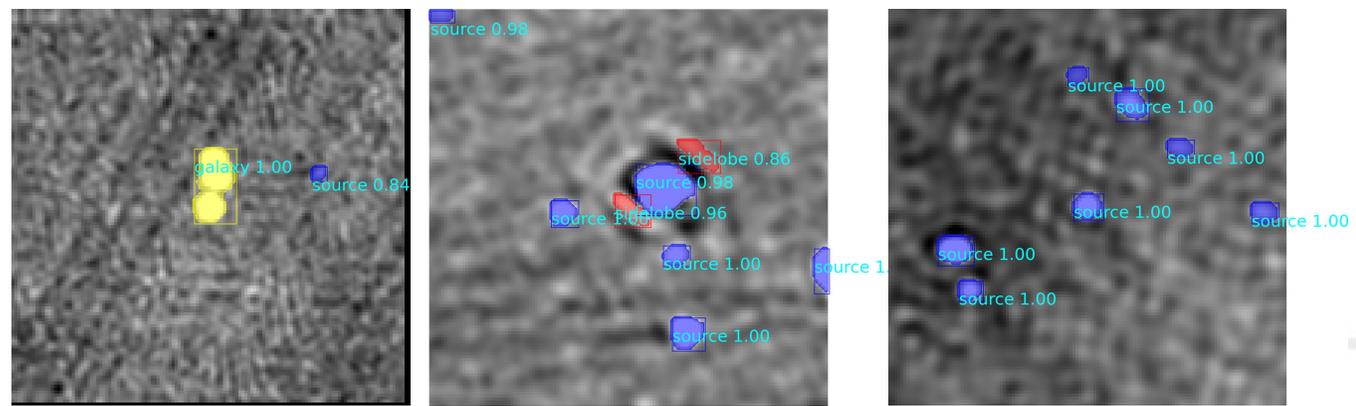
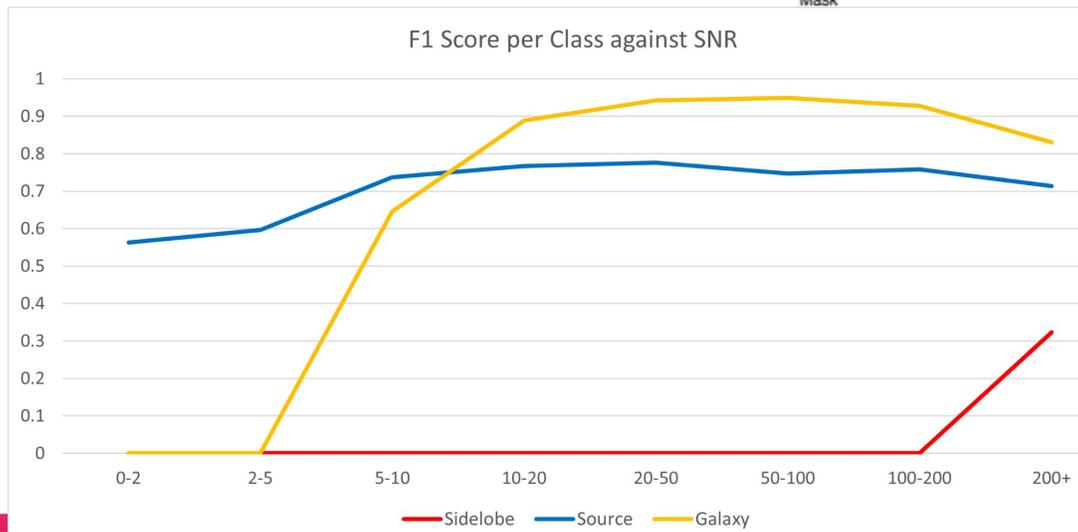
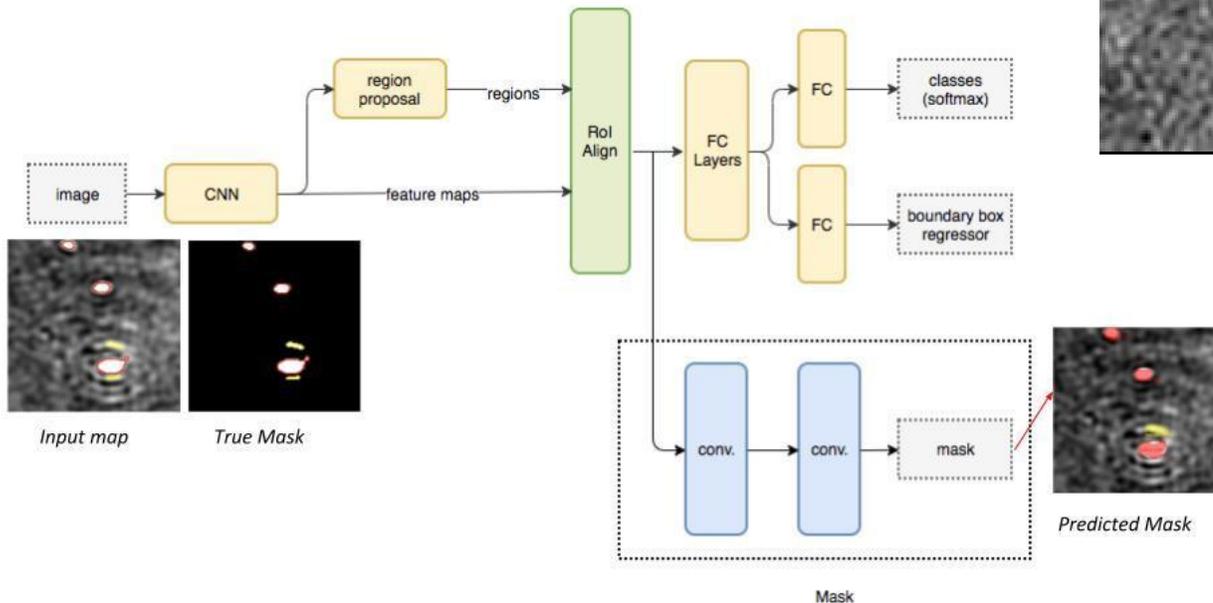
Servizio web REST-ful basato sul framework **Flask** per la ricerca automatica di sorgenti con il software **CAESAR**

- Strategie di **esecuzione multipla** su
 - cluster GARR Kubernetes e
 - INAF/CIRASA Slurm
 - processamento asincrono con celery
- Accessibile da **interfaccia utente web** (Django)
- Altri source finder (ASGARD, CUTEX, Aegean) in fase di integrazione
- Integrazione con il client di visualizzazione ViaLactea in corso

Maggiori dettagli @ Riggi, S., et al. (2021)
Astronomy and Computing
<https://github.com/SKA-INAF/caesar-rest>



ML/DL Analysis



- Basato sul framework di **object detection** (Mask R-CNN) e sul framework di **segmentazione semantica** (modello U-Net)
- Addestrato a riconoscere **sorgenti puntiformi, radiogalassie estese e artefatti di imaging** (sidelobes) su dati ASKAP, RGZ FIRST, ATCA e MeerKAT (~ 15k immagini)
- Utilizzato come source finder standalone (detection + classificazione) o come classificatore per altri finders (es. Caesar)
- Test su **piattaforme multi GPU** come NVIDIA DGX @ CINECA
- **Versione parallela MPI** in versione alpha
- Studi di **ottimizzazione delle prestazioni** in corso

Maggiori dettagli @ D. Magro, et al. (submitted) *PASA*,
C. Pino, et al. (2021) Springer
<https://github.com/SKA-INAF/mrcnn>

Conclusioni

Produzione requisiti della rete degli SRC “Federated Computing Services (WG3)”

- Sincronizzazione e omogeneizzazione dei requisiti con gli altri SRC WG e in particolare:
 - SRC Network Architecture (WG0)
 - Science-Archive VO FAIR (WG4)
 - Science User Engagement (WG6)

Prototipazioni @ GARR Cloud

- Servizi di federazione operativi su Cloud
- Servizi di ricerca delle sorgenti che includono anche tecniche avanzate di ML sviluppati, integrati e distribuiti su EOSC marketplace
- Servizi di Visual Analytics integrati e distribuiti su EOSC marketplace

Prossimi passi

- Finalizzazione requisiti
- Sviluppo primi prototipi operativi dello SRC: AAI, distribuzione dati, ambiente notebook, visualizzazione.
- Omogeneizzazione API verso interoperabilità SRCNet

Grazie!

Collaboratori: S. Riggi (INAF), F. Vitello (INAF), G. Tudisco (INAF), C. Bordiu (INAF), F. Bufano (INAF), C. Pino (INAF), U. Becciani (INAF), D. Magro (UoM/INAF), R. Sortino (UNICT/INAF), A. De Marco (UoM), C. Spampinato (UNICT), K. Zarb Adami (UoM), G. Fiameni (NVIDIA), C. Pisa (GARR), M. Lorini (GARR)

INAF-OACT IT & Radio Astronomia

WORK
SHOP
GARR
2021

NET
MAKERS



 **Mentimeter**

9489 0567