
ASTRON
Netherlands Institute for Radio Astronomy

NL SRC Community Meeting

- Welcome!
- Zoom: https://edu.nl/8hjp4
  - Please connect to Zoom & share your screen to display slides.
  - ...but keep your mic muted & turn the volume down!
- Rough programme:
  - 13:00–14:00: Plans for & status of SKA and SRC construction
  - 14:00–15:20: Science use cases (& coffee break)
  - 15:20–16:00: Discussion, Q&A, next steps
  - 16:00: Borrel

SRC Design & Construction

NL SRC Community Meeting

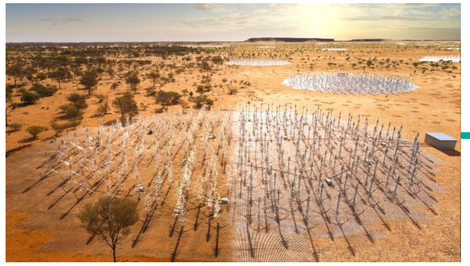
John Swinbank*, swinbank@astron.nl

ASTRON
Netherlands Institute for Radio Astronomy

*Thanks to Rosie Bolton & Jesús Salgado (both SKAO) for various images and slides

- Why are we here?
- Introductions!
  - *Michiel:* NL SKA Office, SRC Steering Committee Chair
  - *John:* ASTRON SDC Programme Manager
  - Others as we go...

What is a regional centre?



SKA Central Signal Processor

SKA Science Data Processor

SKA Central Signal Processor

SKA Science Data Processor

350 PB/year

Observatory Data Products

350 PB/year

SKA Regional Centre

SKA Regional Centre

SKA Regional Centre

SKA Regional Centre

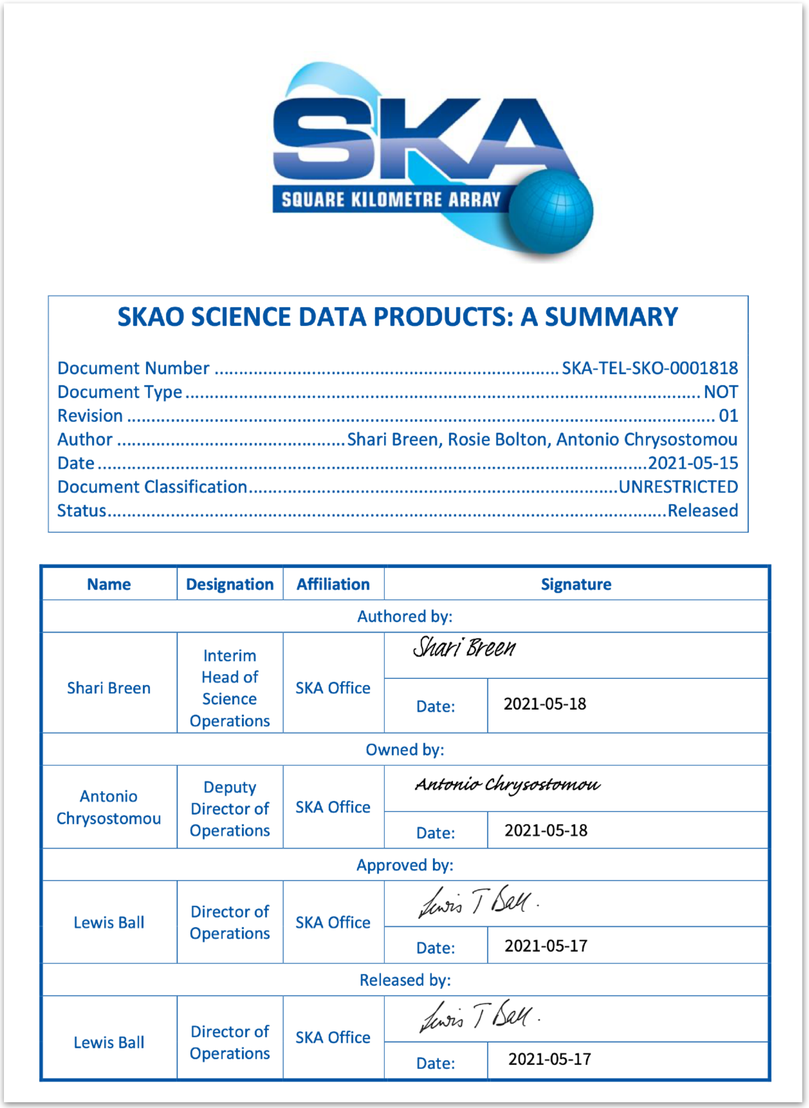
...

- The SKA Observatory will generate “science ready” data (images, etc).
  - Science users have no direct access to the processing system or data products; it is operated by SKAO staff only.

- Data products are supplied by SKAO to a network of *SKA Regional Centres.* SRCs are responsible for:
  - Archiving and curating science data;
  - Making that data available to end users;
  - Enabling scientific discovery.

Data products

- The SKA Observatory will produce *Observatory Data Products* (ODPs) and provide them to the Regional Centre network.
- The ODPs are defined in SKAO documentation, e.g. SKA-TEL-SKO-0001818 (👉), and will include:
  - Image cubes
  - *uv* grids
  - Calibrated visibilities
  - Local sky model catalogue
  - Imaging transient source catalogue
  - Pulsar timing solutions
  - Transient buffer data
  - Sieved pulsar & transient candidates
  - Science alerts catalogue
- The ODPs are augmented by *Advanced Data Products* (ADPs) which are generated within the regional centres.



SKAO SCIENCE DATA PRODUCTS: A SUMMARY

Document Number SKA-TEL-SKO-0001818
Document Type ..... NOT
Revision 01
Author ..... Shari Breen, Rosie Bolton, Antonio Chrysostomou
Date 2021-05-15
Document Classification ..... UNRESTRICTED
Status ..... Released

Name	Designation	Affiliation	Signature	
Authored by:				
Shari Breen	Interim Head of Science Operations	SKA Office	Shari Breen	
			Date:	2021-05-18
Owned by:				
Antonio Chrysostomou	Deputy Director of Operations	SKA Office	Antonio Chrysostomou	
			Date:	2021-05-18
Approved by:				
Lewis Ball	Director of Operations	SKA Office	Lewis T Ball.	
			Date:	2021-05-17
Released by:				
Lewis Ball	Director of Operations	SKA Office	Lewis T Ball.	
			Date:	2021-05-17

https://bit.ly/SKAO-ScienceDataProducts

SRC capabilities

Scientific Applications
Analysis tools, notebooks, workflow execution, machine learning, ...

Distributed Data Processing
Computing capabilities provided by the SRCNet to allow data processing

Data Discovery
Discovery of SKA data from the SRCNet, local or remote, transparently to the user

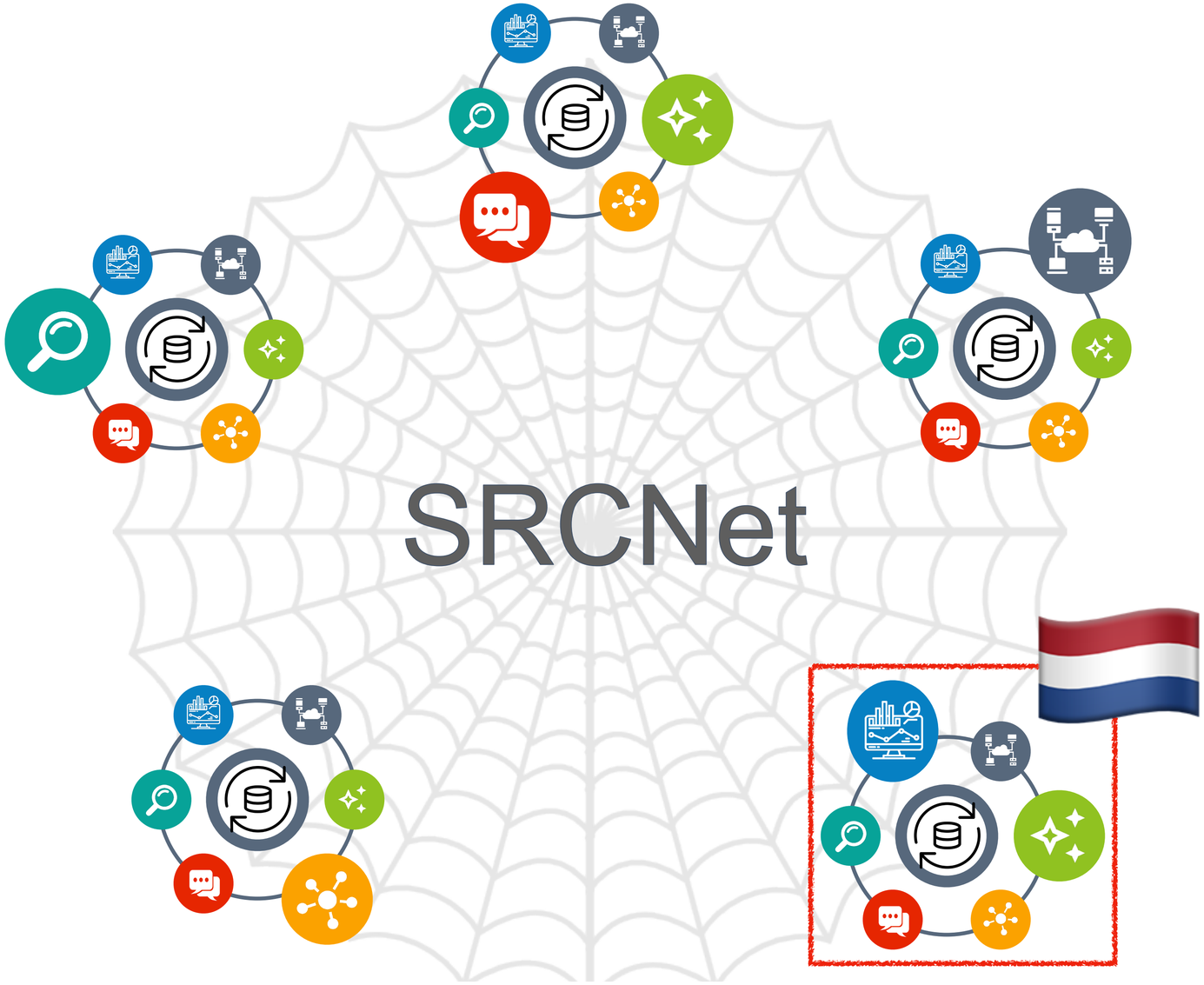
Visualization
Advanced visualization tools for both SKA data and data from other facilities

Support to Science Community
Support community on SKA data use, SRC services use, training, project impact dissemination

Interoperability
Heterogeneous data and services from various SRCs and other facilities

Data Management
Dissemination of data to SRCs and distributed data storage

Global SRC network



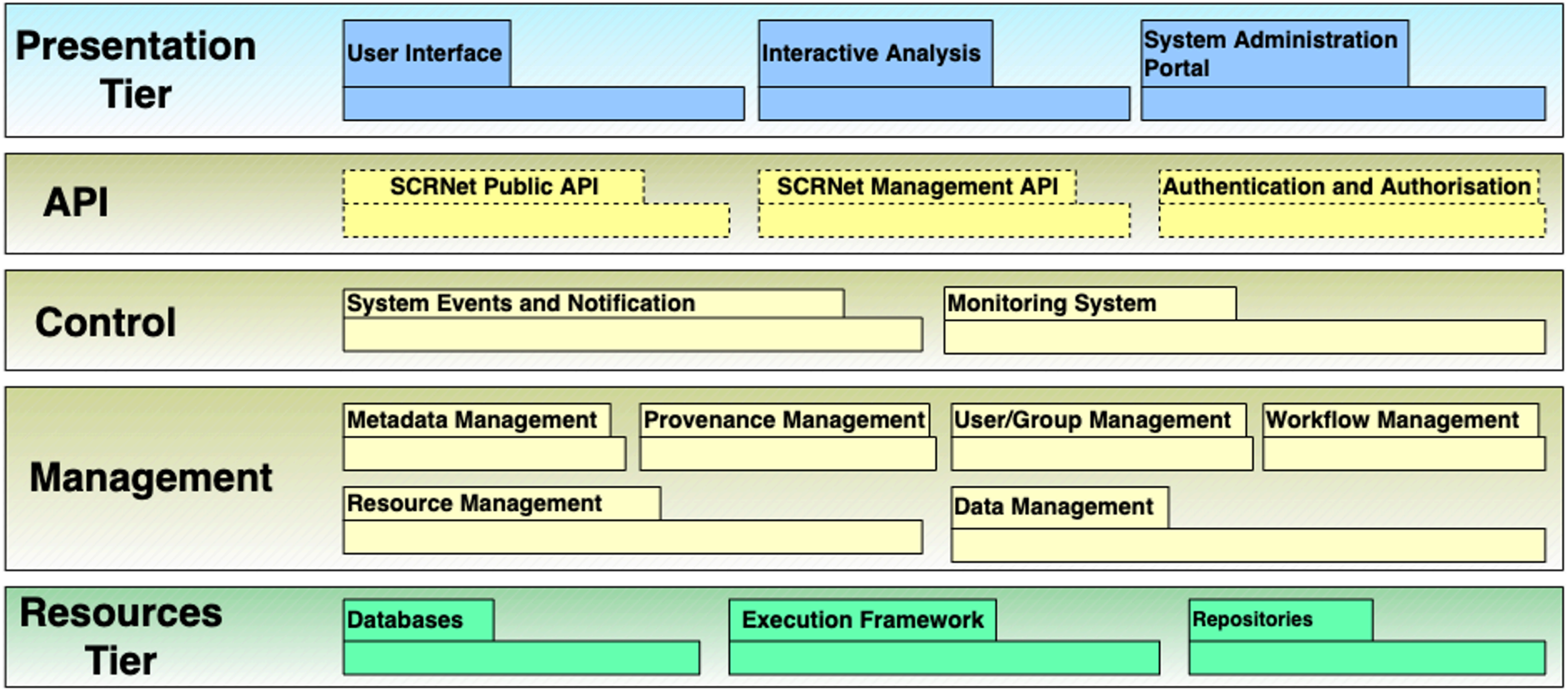
Collectively meet the needs of the global community of SKA users.

Heterogeneous SRCs, with different strengths, working together in a federated infrastructure.

SRC design “principles”

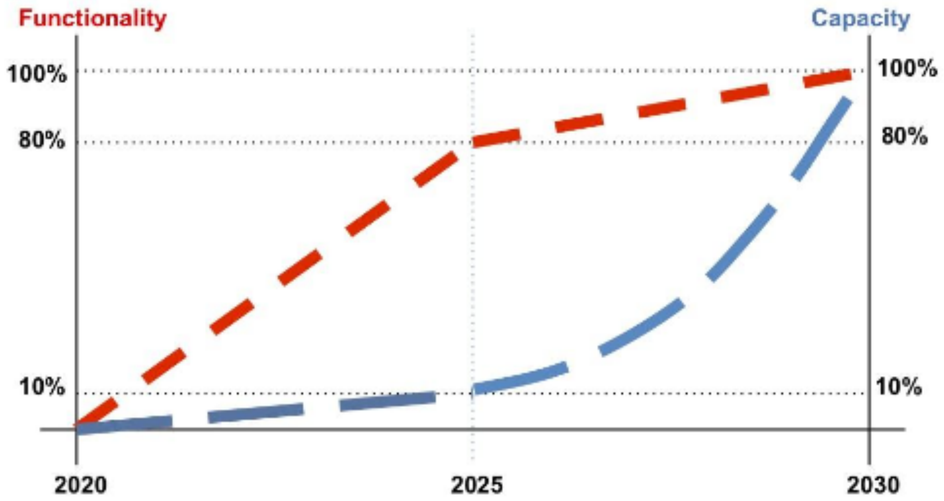
- Global effort, with all SRCs contributing, but geographically transparent to the end user
- Scalable, both globally and locally
- Extensible to add new capabilities
- Resilient, against data loss and service disruption
- FAIR (Findable, Accessible, Interoperable, Reusable)
- Reproducible: any analysis or data product can be reproduced at a later time
- Usable and accessible, to folks with a wide range of experience levels and/or physical needs
- Re-use and extend existing tools where possible
- Distinguish *policy* from *implementation*

SRCNet Architecture



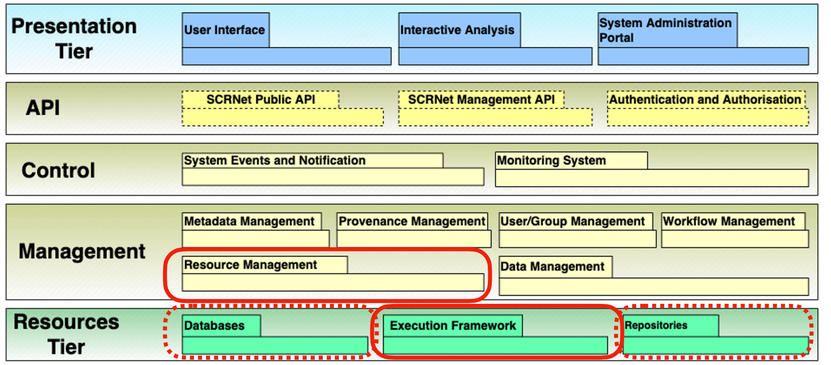
Development timeline

2021	Initial requirements collection
2022	**Prototyping (now!); 50+ people involved**
2023	Agreement of implementation plan; start of construction
2024 onwards	Limited operations in support of construction & commissioning
2025 onwards	Scale up to meet full SKA capacity



Federated & Cloud Infrastructure

- The SRC network will span multiple different providers and technologies.
  - Cloud based, physical infrastructure, different types of cluster, HPC/HTC systems, middleware, etc.
- A “spanning layer” will abstract away the underlying details so that user services — and the users themselves — are shielded from complexity.
- Making this work efficiently is a real challenge!
- Current activities focus on developing a common understanding and exploring existing solutions.



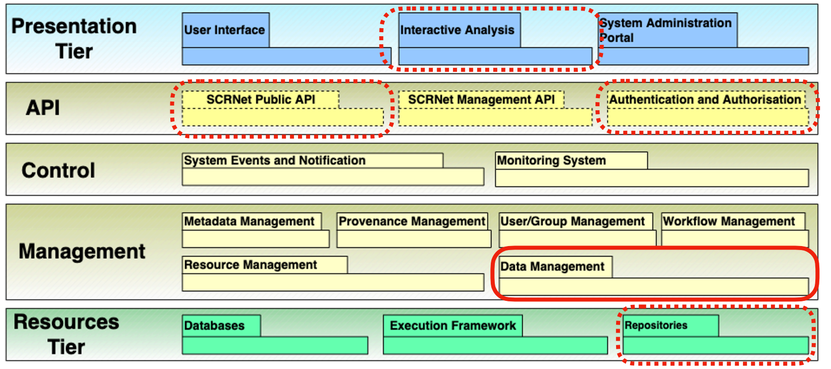
Data Logistics

- No single SRC will be able to store or process all of the data.
- Together, the SRCs will cooperate to:
  - Receive data from the telescope;
  - Ensure it is archived securely and with an appropriate level of redundancy;
  - Make data available where & when it is required for processing.
- *Policy-based* data management solutions can help us automate this, rather than relying on human operators.
- The system needs to be *smart* and *transparent*: users shouldn't be kept waiting for their data, and data transport should be minimized (for environmental & cost reasons).
- Current activities include building out a Rucio test network and integrating it with astronomical data standards (e.g. IVOA).

SCIENTIFIC DATA MANAGEMENT

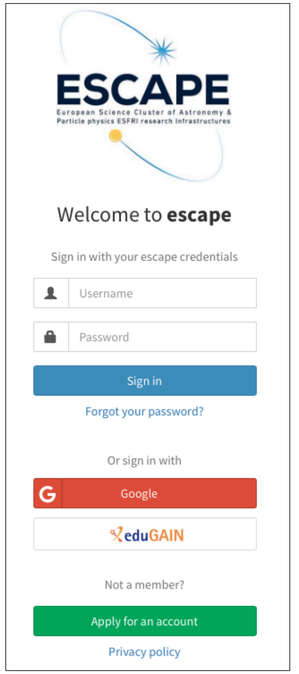
RUCIO

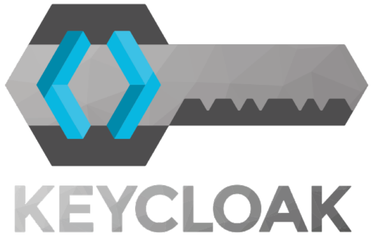
CANFAR

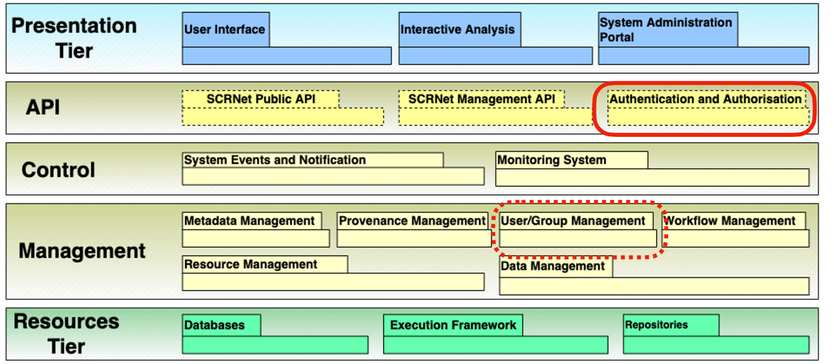


Authentication & Authorization

- Make sure that people have access to appropriate data, resources, etc.
- In a scaleable way — we can't maintain a single database of all SKA users worldwide, so we delegate to local “identity providers”.
- There's a blizzard of existing acronyms and technology to wade through here — IAM, AARC, OAuth, OIDC, X.509, LDAP.
- Current activities focus on building consensus among the various national SRC projects and infrastructure providers.

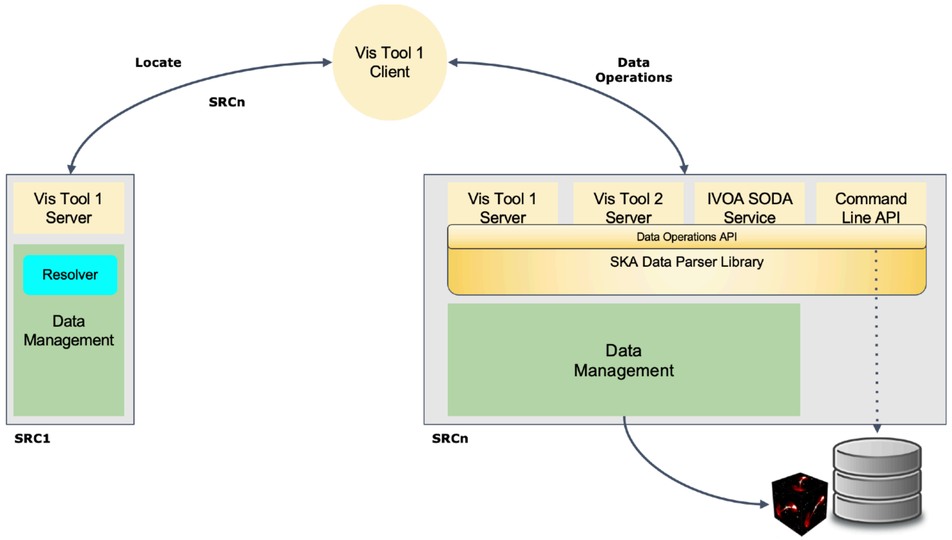


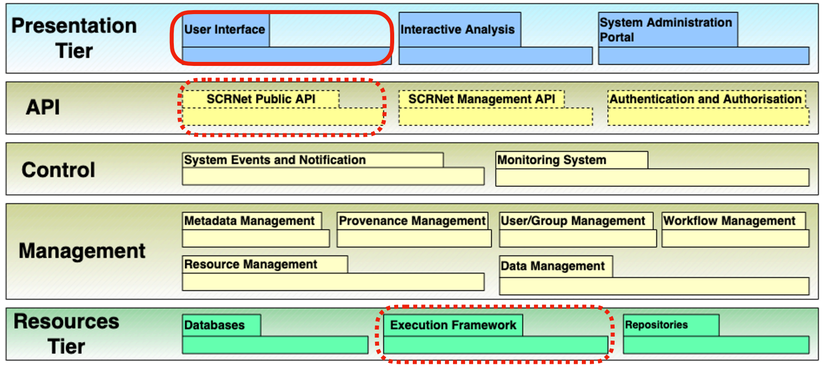




Data Visualization

- How can we effectively visualize massive SKA datasets...
- ...without transferring enormous quantities of data?
- Modern on-line visualization tools support “smart” data selection, only transferring the data that you are actually looking at.
- Unable / unlikely to develop completely new tools for SKA.
- Aim to provide standard data selection interfaces, then adapt and modernize existing tooling to make use of them.





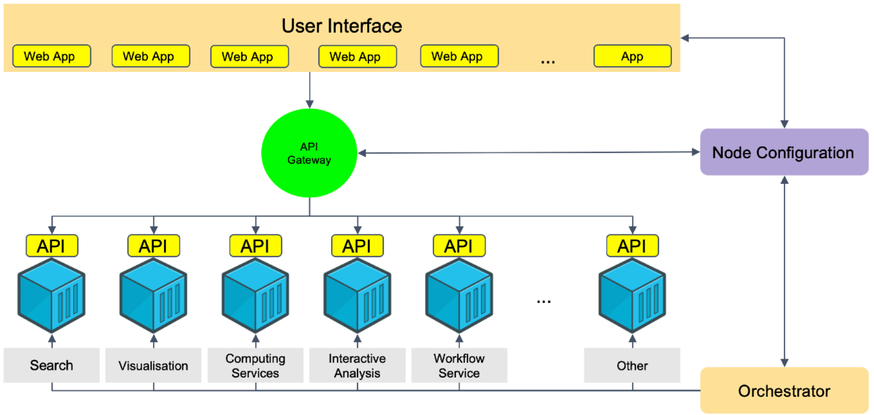
Science Platform

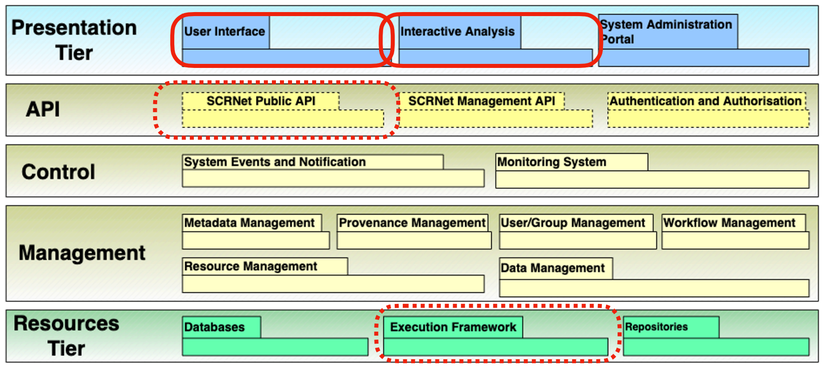
ESCAPE ESAP | ESFRI Science Analysis Platform

jupyter

CANFAR

- Your primary interface to the SRC Network — where the science user will come to discover, access, analyze, and publish data.
- Like visualization tooling, this is primarily a matter of selecting, adapting, and integrating existing tools.
- Aim to provide familiar interfaces (e.g. Jupyter notebooks, but also traditional command line and GUI applications).
- *This is where the ASTRON team is currently most heavily involved.*
- Current focus is on a competitive evaluation of promising platform technologies.

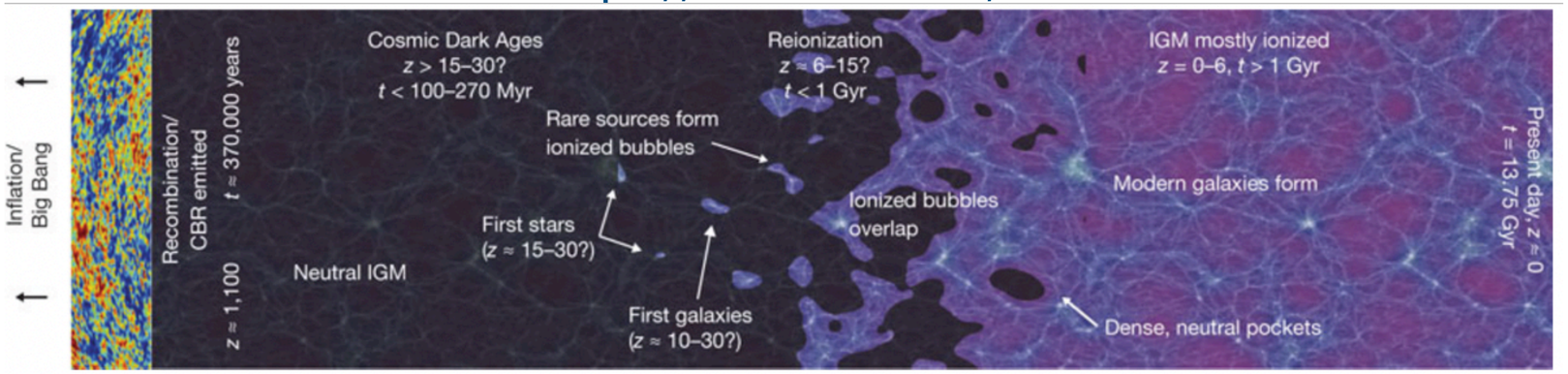




Science Data Challenge 3

- “Prepare the radio-astronomical community for the novel nature of the data expected from the Square Kilometre Array”
- Participants of this data challenge will be tasked with elucidating exactly when (given a realistic, artificial dataset) the Epoch of Reionization occurred. However, given that the observation itself is extremely challenging, our challenge will be broken down in to two parts:
  - Removal of foreground emission from Galactic and Extragalactic sources ('Foregrounds', or SDC3a)
  - Inference of important parameters of the Epoch of Reionization ('Inference', or SDC3b)”
- ASTRON/SURF providing computing & storage resources — as a scientific community, please engage!
- Increasing evolution of science data challenges towards SRC construction and operations with time.

https://sdc3.skao.int/



Robertson et al. (2010)

Science Applications

- The prototypes focus on infrastructural services and software.
- ASTRON also has the ambition to support the development of science applications or pipelines which will run within the SRC network.
- *This is where you come in*. We will look to develop collaborations between the development teams at ASTRON (& SURF, if appropriate) and Netherlands research community.
- This could include e.g.
  - Adapting and extending existing tools to access data through the SRC “spanning layer”.
  - Development of new tools that explicitly target SKA-sized data.
  - *...your ideas here.*