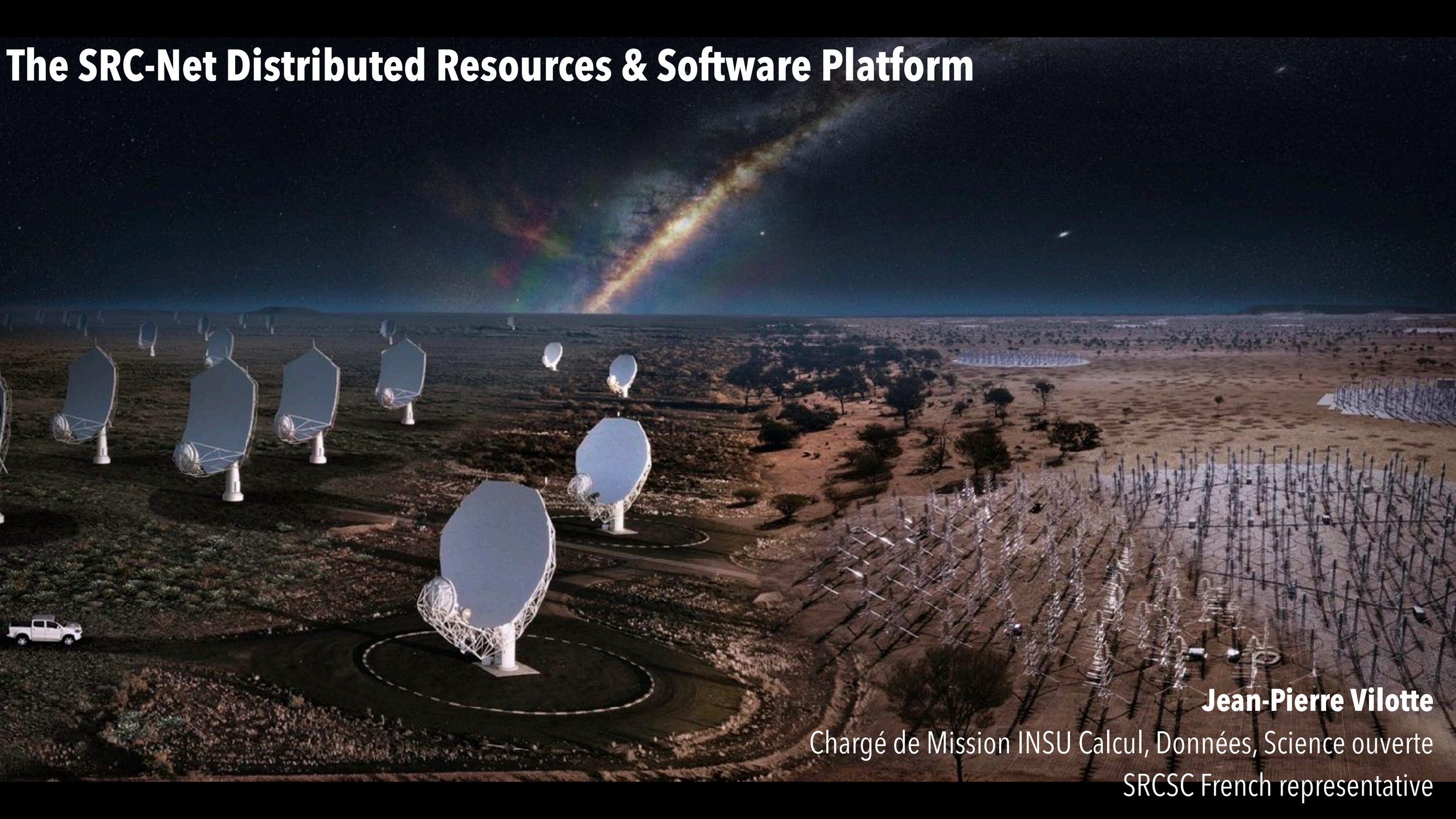
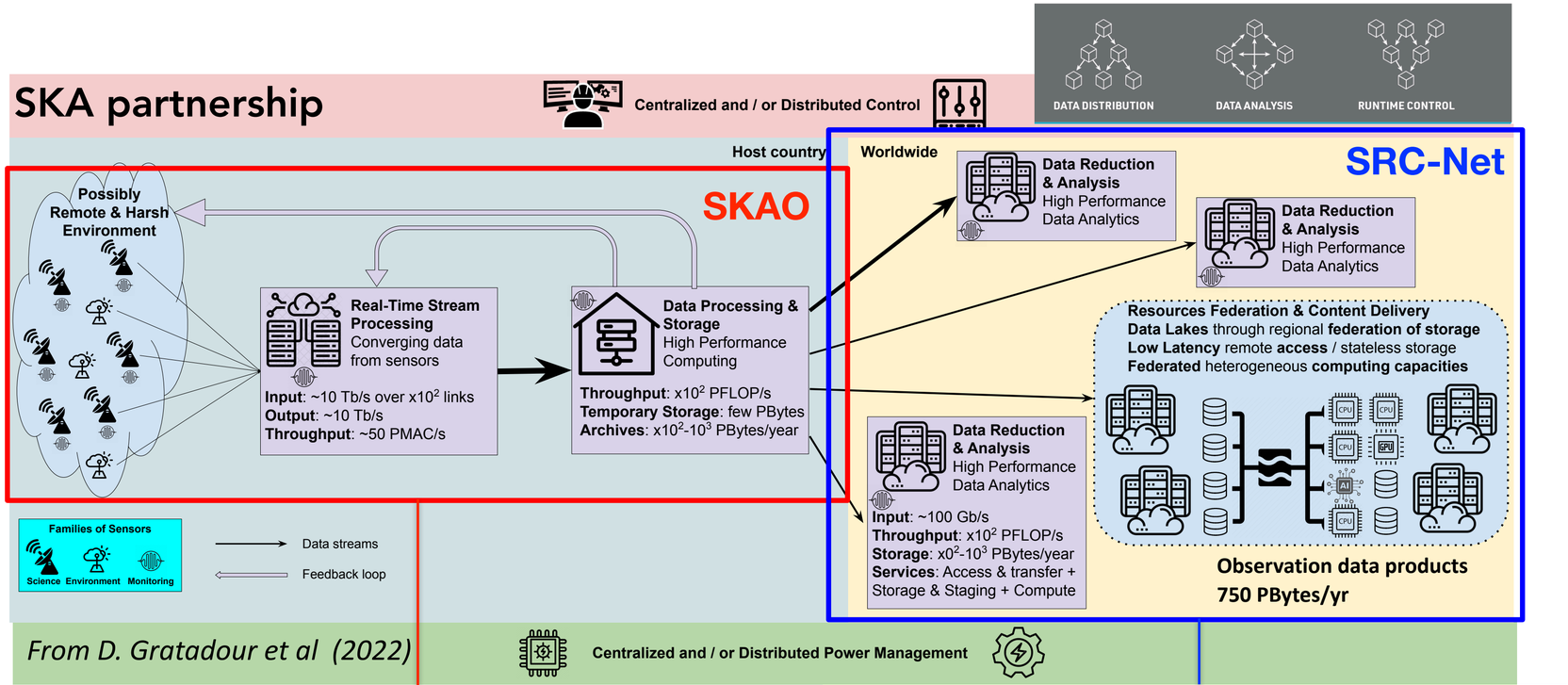
The SRC-Net Distributed Resources & Software Platform



Jean-Pierre Vilotte
Chargé de Mission INSU Calcul, Données, Science ouverte
SRCSC French representative

SRC-Net: an SKAO end-to-end partnership



Observation data product delivery
- Data streaming reduction and processing,
- Edge computing and content delivery network
- Centralised HPC infrastructures (SDP)

Distributed Data archive and Data analysis
- Emerging hybrid Cloud technologies: compute, data logistics, storage, wide-area workflows
- Federated distributed & heterogeneous resources
- logistical networking and Power management

Advanced science-driven SRC-Net capabilities to: bridge geographic, organisational, and technological boundaries; foster collaborations across SKA science communities and a end-to-end partnership with SKAO.

No access to the SDPs nor to the raw SKAO data

The SRC-Net: a critical component of SKA

Federation of Resources to fully process, archive, curate and scientifically use the SKA observation data products

Beyond the scope of the SKA1 construction and operation budget

In March 2016 the SKAO Data Flow Advisory Panel recommended that the SKAO Board encourage SKAO member states to form **a collaborative network of SKA Regional Centres (SRCs) to provide the essential functions not presently funded within the scope of the SKA1 project**.

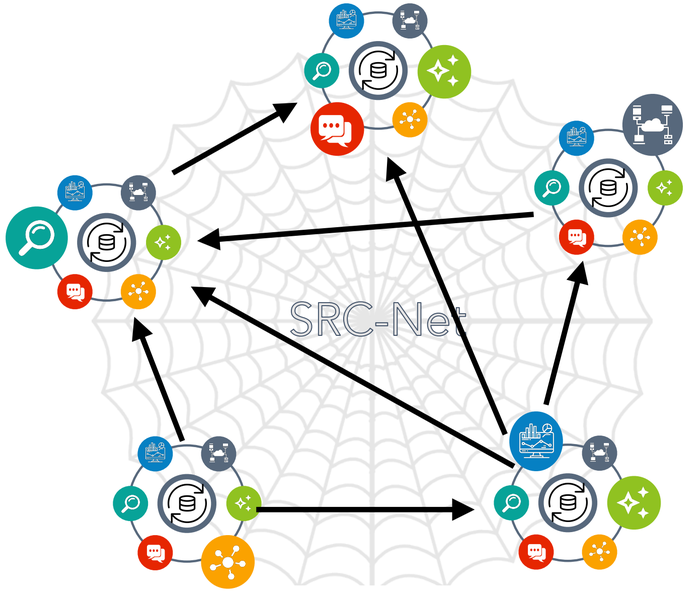
SKAO and the international SKA science community need to work collaboratively to shape and establish the shared and distributed SRC-Net data, computing and networking capacity and capability.

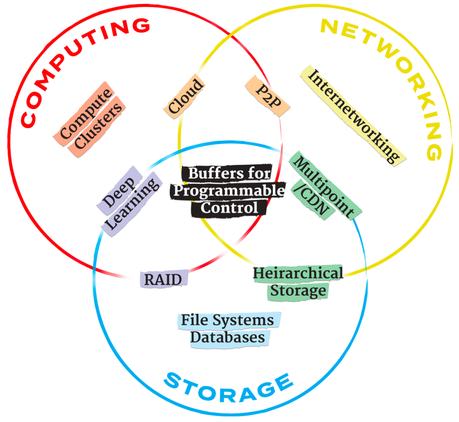
SRC-Net must act like a single logistically coherent and operational system that is **persistent**.

Only possible through international cooperation and collaboration: resources pledged from multiple governments and stakeholders, with a variety of funding tools

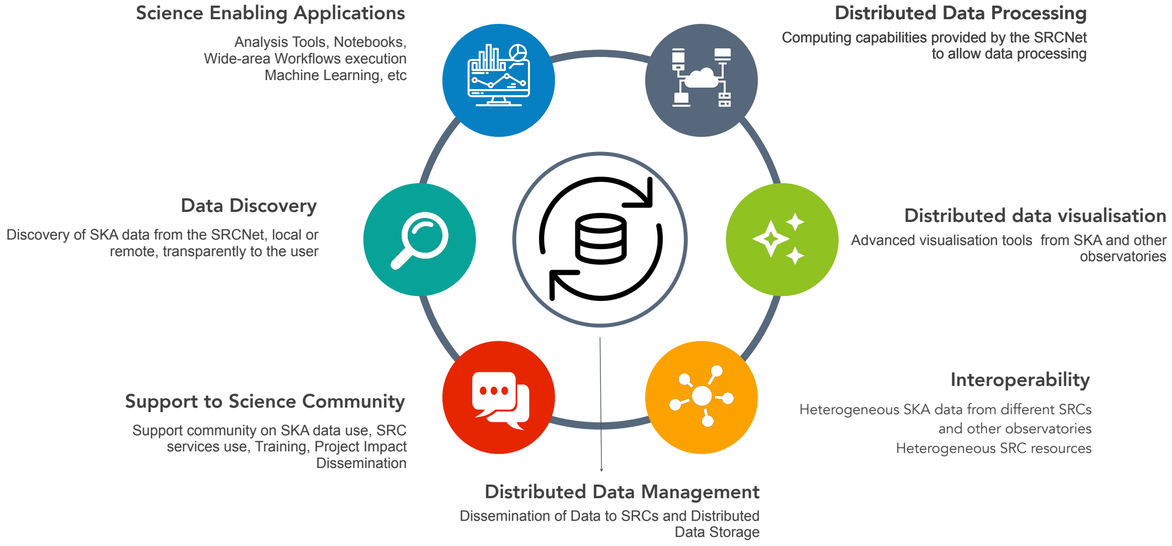
Transformational research practices and technological developments: require **major collaborative and interdisciplinary efforts** across algorithmic research, software development and integration, data logistics, Infrastructures providers and science application teams.

SRC-Net: Science Distributed Software Platform & Federation of Resources





Ensure that scientists can access SKA data products and use them to make discoveries



SRC-Net

Governance and policy: bridging geographic, political, organisational, and technological boundaries

Heterogeneity and isolation: parts of a life in globally distributed environment, modelled by the SRC-Net architecture and managed by the SRC-Net policy

Cybersecurity and sustainability: at the system level, hyper-vision of the resources topology used in workflows deployment

Architecture: enable scalability deployment of a wide diversity of science-driven workflows blurring lines between isolated silos (storage, networking, processing) together with a metadata centric approach

Distributed FAIR data management: massive and diverse SKA observation and science data products, workflows, software librairies …

Science-driven platform: start with science-dependent workflows and their data collections, diverse patterns of where and how data are accessed, transformed, and intermediate results managed

Interoperability/Portability: internet-based software stack providing networking, storage, processing as services through shared protocols despite differences in implementation

Open source and converged standards: foster broad consensus, transparency and adoption by the SKA science community

Shaping strategy: incentives to foster voluntary adoption by all stakeholders far more desirable that is legislated or imposed by a legal organisation

SRC-Net: Data products archiving and management

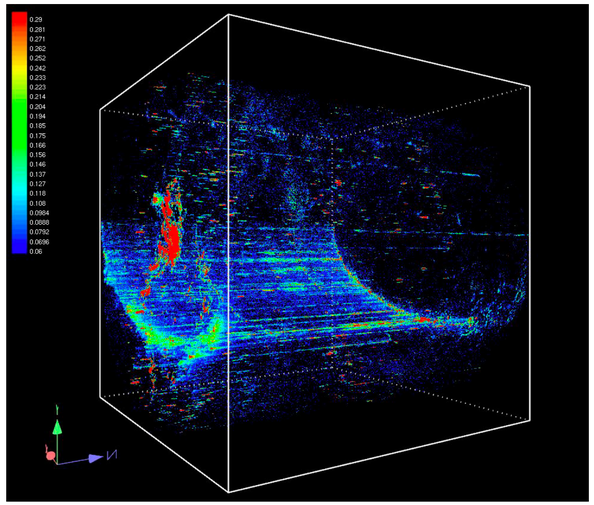
FAIR SKA data products: open access (in time)

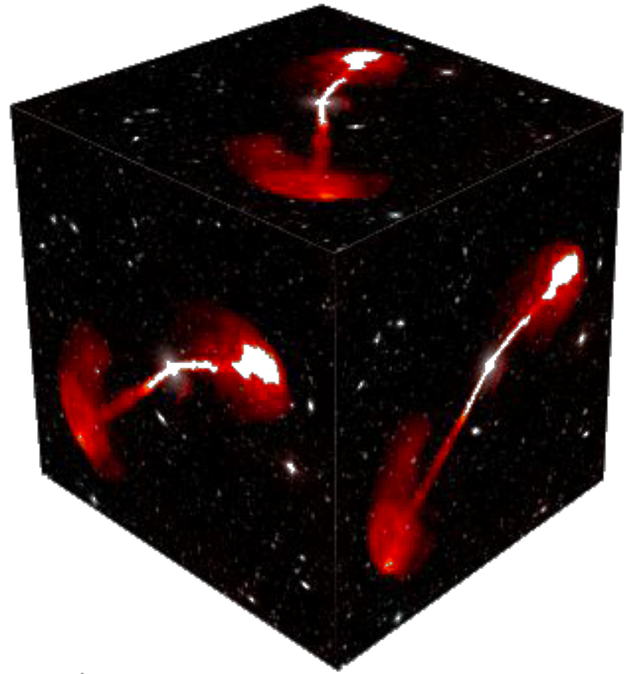
SKA science archive build
- around IVOA standards
- interoperable with other archives from reference experiments

Storing SKA observation and user-generated advanced data products is a challenge
- observation data products (700 PB/yr in phase I)
- different product-dependent data layout
- several million dollars per year in new data, for one copy

Globally distributed data management to
- enable best possible use of pledged storage and communication resources,
- avoid unnecessary duplication
- support mirroring of popular data products to enhance user experiments

Data logistics across SRC Net should
- support federated execution of a variety of wide-area workflows (HPC, HDA, AI)
- support exploring and visualising distributed large data objects
- enable best possible use of heterogeneous SRC-Net resources





SRC-Net: Global collaborative capabilities

Collectively meet the needs of the SKA users community

Federate heterogeneous SRC nodes with different policies, capabilities and expertise

SRC nodes pledge resources into the global Federation of Resources, and support SRC-Net services and operation

Users access SRC-Net resources according to their research needs and permissions (AAI)

Contribution effort from each national SRC nodes is expected to be proportional to their SKAO fraction

Additional resources at an SRC node could be given to the pool or prioritised to support national interests

SRC Operation Group (SOG) is composed of identified personnel and expertise within each SRC

SOG - SKAO partnership with a team from across each SRC and SKAO Ops

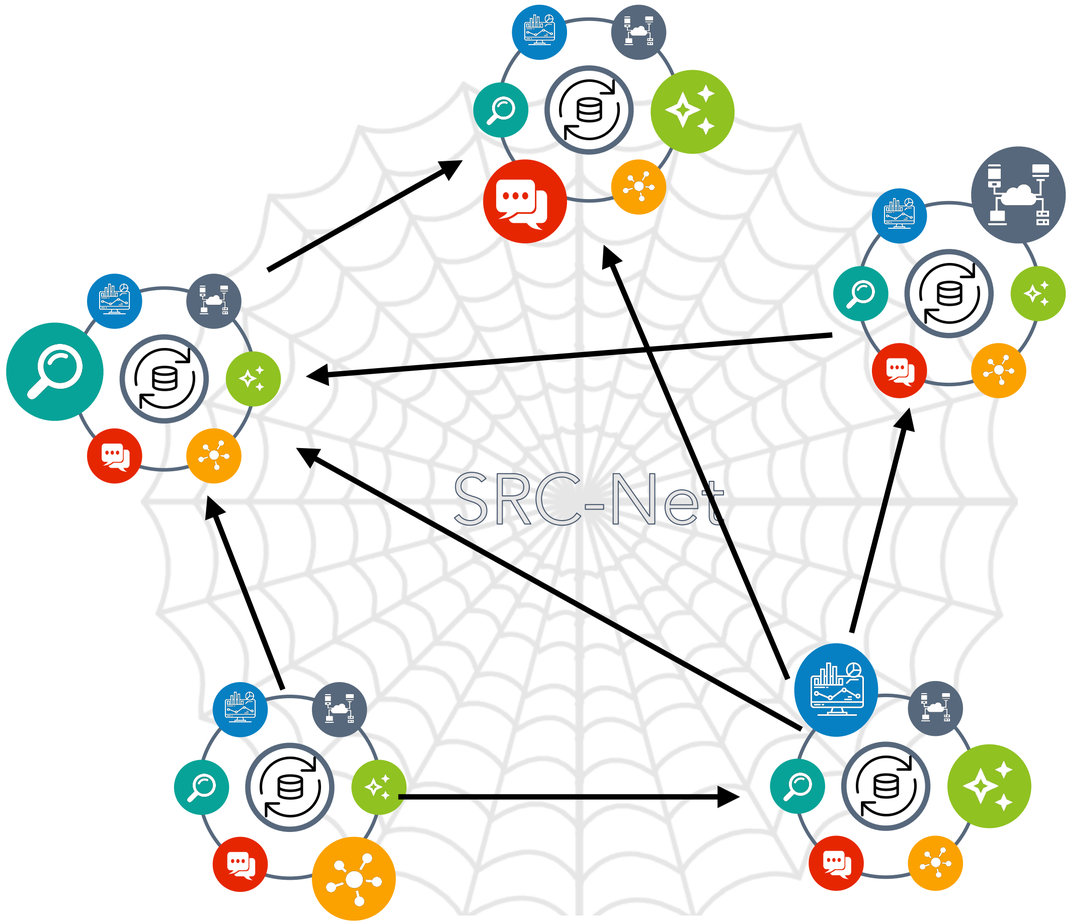
Energy: an overarching challenge for SRC-Net sustainability

- SKA data movement to archival and curation SRC-Net nodes, to each scientific step that accesses the SKA data, to visualisations, **all represent a significant use of energy.**

- **The SRC-Net , which spans autonomous national/regional resources, needs to minimise energy consumption at the system level to reduce environmental impact and maximise resilience to energy volatility**

- **Steps toward energy minimisation:**
  - reduce computing costs by enabling the use of SRC-Net components well-matched to the stage within the scientific method;
  - reduce data-movement costs by enabling caching, collocation of data and compute resources in a locality, and data compression;
  - enable re-use of processing and science data through effective sharing, metadata, and catalogs – a strategy that a provenance system supports well, and
  - reduce computing system entropy (e.g. workload interference, system jitter, tail latency, etc) through on-demand isolation, noise-resistant priority, cache QoS, and novel uncertainty bounding techniques.

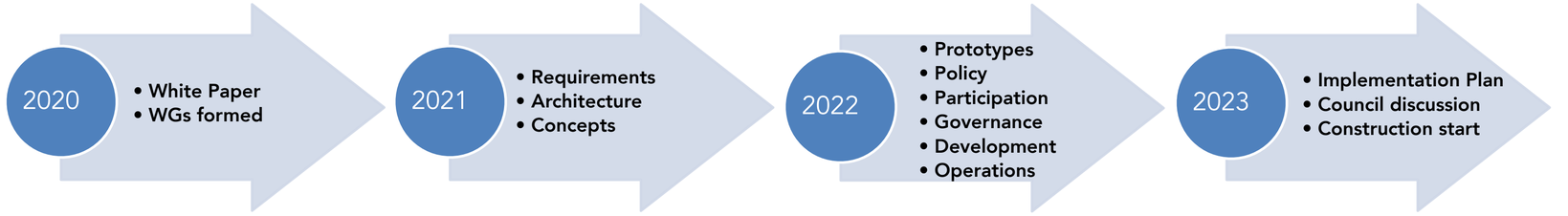
- **Governance, policy and accounting model is needed** to optimise energy savings and to sustain the respect for the value of the contributing stakeholders.



SRC Steering Committee

SRC Steering Committee

Michiel. van Haarlem (Netherlands, Chair))
Karen Lee.-Waddell (Australia),
Bradley Frank (South Africa),
Tao An (China),
Severin Gaudet (Canada),
Yogesh Wadadekar (India),
John Conway (Sweden),
Hans-Rainer Klockner (Germany),
Domingos Barbosa (Portugal),
Andrea Possenti (Italy),
Lourdes Verdes-Montenegro (Spain),
Emma Tolley (Switzerland),
Jeremy Yates (UK),
Jean-Pierre Vilotte (France),
Antonio Chrysostomou (SKAO),
Takuya Akahori (Japan, observer),
Hyunwoo Kang (Korea, observer)
André Csillaghy (SEAC, Observer)
Rosie Bolton (SKAO, invited member)



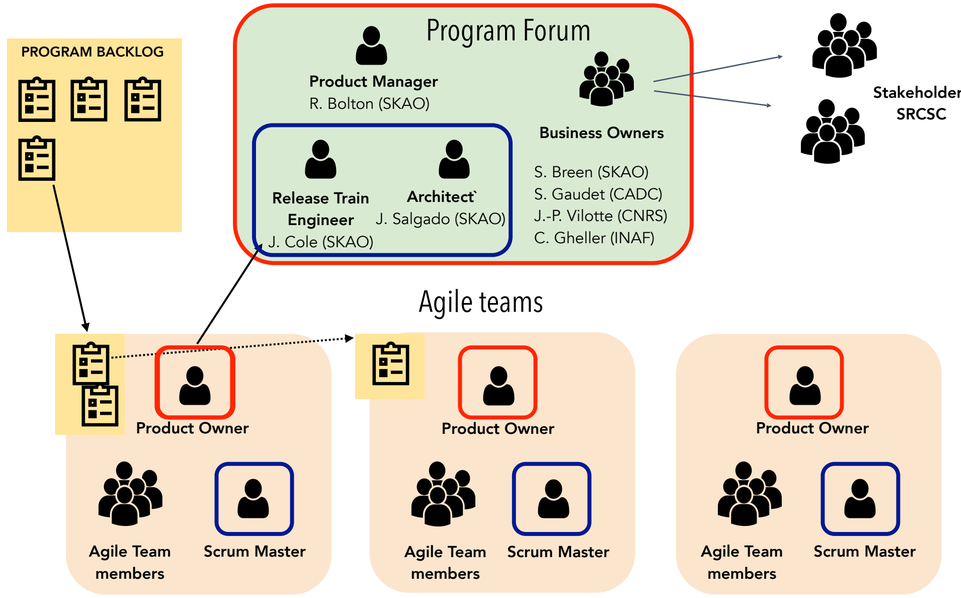
From Working Groups to identify first level requirements

WG0 - SRC Network architecture: ***J. Salgado (SKAO);*** **WG1 - SRC-Net Data Management and Logistics:** ***Rosie Bolton*** **(SKAO) &** ***An Tao*** **(China); SRC-Net operations:** ***A. Chrysostomou*** **(SKAO) &** ***M. van Haarlem*** **(Netherlands); WG3 - SRC Net distributed software platform:** ***J.-P. Vilotte*** **(France) &** ***Y. Wadadekar*** **(India); WG4 - SRC-Net SKA FAIR science archive:** ***L. Verdes-Montenegro*** **(Spain) &** ***S. Gaudet*** **(Canada); WG5 - SRC-Net hybrid Cloud/HPC resources:** ***J. Yates*** **(UK) &** ***D. Barbosa*** **(Portugal); WG6 - SKA science user engagement:** ***A. Possenti*** **(Italy) &** ***H.-R. Klockner*** **(Germany)**

to iterative prototyping program increments in a Scaled Agile Framework for

- designing the SRC-Net architecture and functionalities;
- reviewing and selecting technologies for implementation
- Deploying prototyping test-beds
- Ensure coverage is complete and capture dependancies

Face-to-Face SRCSC meeting: Paris, 20-23 November



Prototyping phase (PI-0, PI-1): Agile Teams

Work *now happening*

Distributed Data Management and logistics

- Replication, distribution, synchronisation of data products and location index,
- Prototypes 1a (Data Lake - Rucio) and 1b (Sorage Inventory - CADC/IVOA)
- IVOA- Data Lake (Rucio) integration, SDC3 data distribution, explore IAM with IVOA
- R. Bolton (SKAO) & P. Dowler (Canada) – Cyan team)

Distributed Data Visualisation and discovery

- Large Data objects, large volume of data, data layout, large number of users
- different tools and services: CANFAR, VisIVO, CARTA, HiPS, Aladin Light, YAFITS/ ARTEMIS
- data interfaces, map tools to visualisation science use cases, select data sets and data formats; tools deployment on the Chinese SRC
- Fabio Vitello & Giuseppe Tudisco (Italy) – Orange Team
- **French contribution:** Mark Allen, F. Bonnarel, Th. Boch, P. Fernique M. Baumann (0.5 FTE, CDS); Ph. Salome, A. Loh; B. Cecconi (0.5 FTE, OP)

Developing and assessing science use cases

- Descriptive use cases backed up by software and data
- Draft living use cases document
- S. Breen (SKAO) & C. Gheller (Italy) - WG6
- **French contribution:** C. Ferrari (CNRS)

Distributed Science Platform

- Science platforms, federated workflow as a service, distributed container management and services, distributed data analysis (notebooks), data logistics
- Evaluation of (30) existing science platforms, Science platform vision document, architecture and scientific assessment of short listed science platforms
- Y. Grange & J. De Boer (Netherland) – Tangerine team
- **French contribution:** J.-P. Vilotte (0.2 FTE, CNRS); B. Cecconi (0.2 FTE, OP/DIO)

Federated Authentication and Authorisation and Identification

- Identification policy and services interoperable with SKAO (Indigo IAM-VOMS-Keycloak), proof of concept of SRC IAM into AUSSRC, integration of IAM with IVOA standards, onboarding - user accounts, groups
- Landscape report, SRC-Net policy definition
- I. Collier (UK) & T. Dack (UK) – Purple team

HPC and Cloud infrastructures and services

- HPC and Cloud technologies and services, storage, networking, processing, programming and execution environments, container images and execution management
- Workshops Aveiro (May 15-16), Cambridge (October 25-26)
- D. Barbosa (Portugal) & XXX (OliveTeam**)**
- **French contribution:** J.-P. Vilotte (0.1 FTE, CNRS)

SRC Net: some issues

Strong current disparity between SRC stakeholders

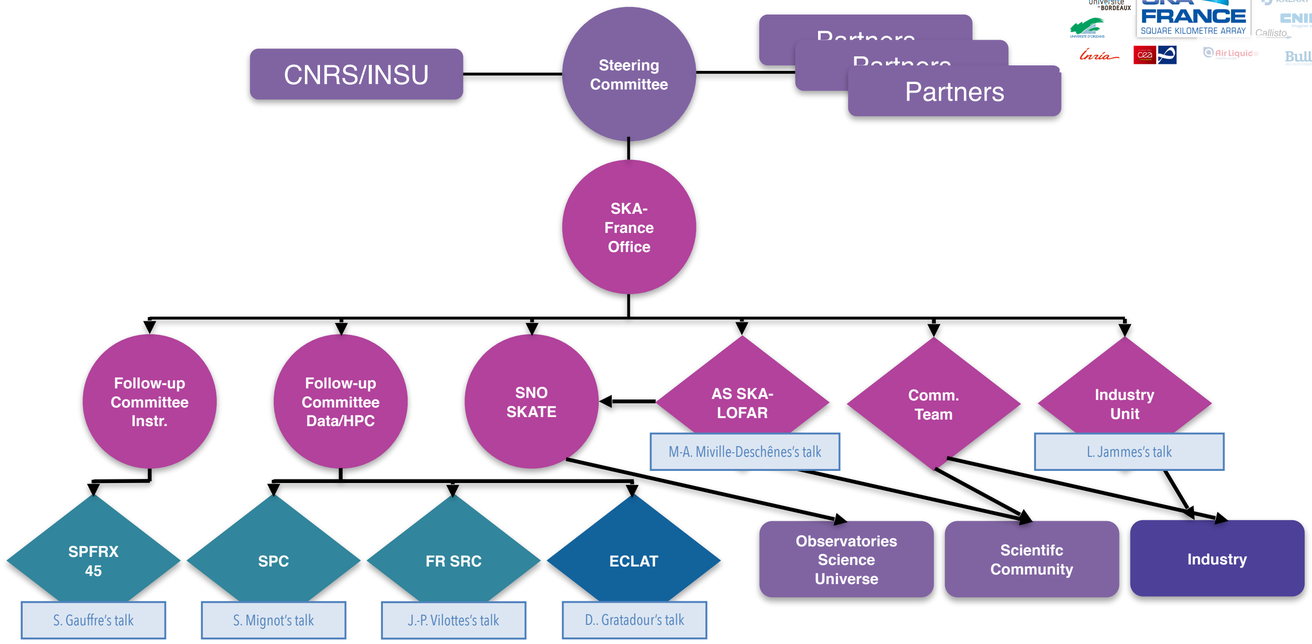
- Some countries have current funding and resources for national SRC project and prototyping activities (Australia, Canada, UK, China, NL, Japan, Switzerland, Germany, Italy, Spain)
  - ➡ They can commit significant resources and expertise
  - ➡ They take the lead in the prototyping activities
  - ➡ They play a strategic role in designing the SRC Net architure

The French current context

- SKA is French Research Infrastructure
- France: SKAO member state with a relatively modest contribution compared to UK, Australia, South Africa, China
- No current funding plane for a French SRC node contribution
- Weak French community engagement in the SRC prototyping activities
- SRC-Net critical for SKA science return and for the sustainability of related pathfinders (e.g., LOFAR/NeNUFAR)

SKA-France: The French SRC-Net contribution

A new collective momentum



SRC-Net Community building

- National science community engagement
- Existing expertise in and synergies with SKA pathfinders (LOFAR/NeNUFAR), IVOA/CDS
- Synergies with other large observation projects in AA (EUCLID, LSST, CTA …)
- Collaborations with INRIA, CEA, GENCI (IDRIS, CINES, TGCC), RENATER
- Synergies with PEPR projects (e.g., ORIGIN, NumPeX, …)
- Synergies with Data Terra and GAIA-DATA

French SRC node contribution to the SRC-Net

- as a distributed science software platform (data management and logistics, processing, visualisation)
- build upon an architecture and a distributed federation of resources (storage, networking, processing)
- leveraging the national landscape: OSUs, regional mesocentres and data centres, national computing and data centres

French contribution to an European SRC-Net contribution

- ESKAF
- European projects (EuroHPC, INFRA Horizon-Europe, EOSC)

Et un grand merci à Michel pour m'avoir accueilli et et guidé dans ce fantastique projet …