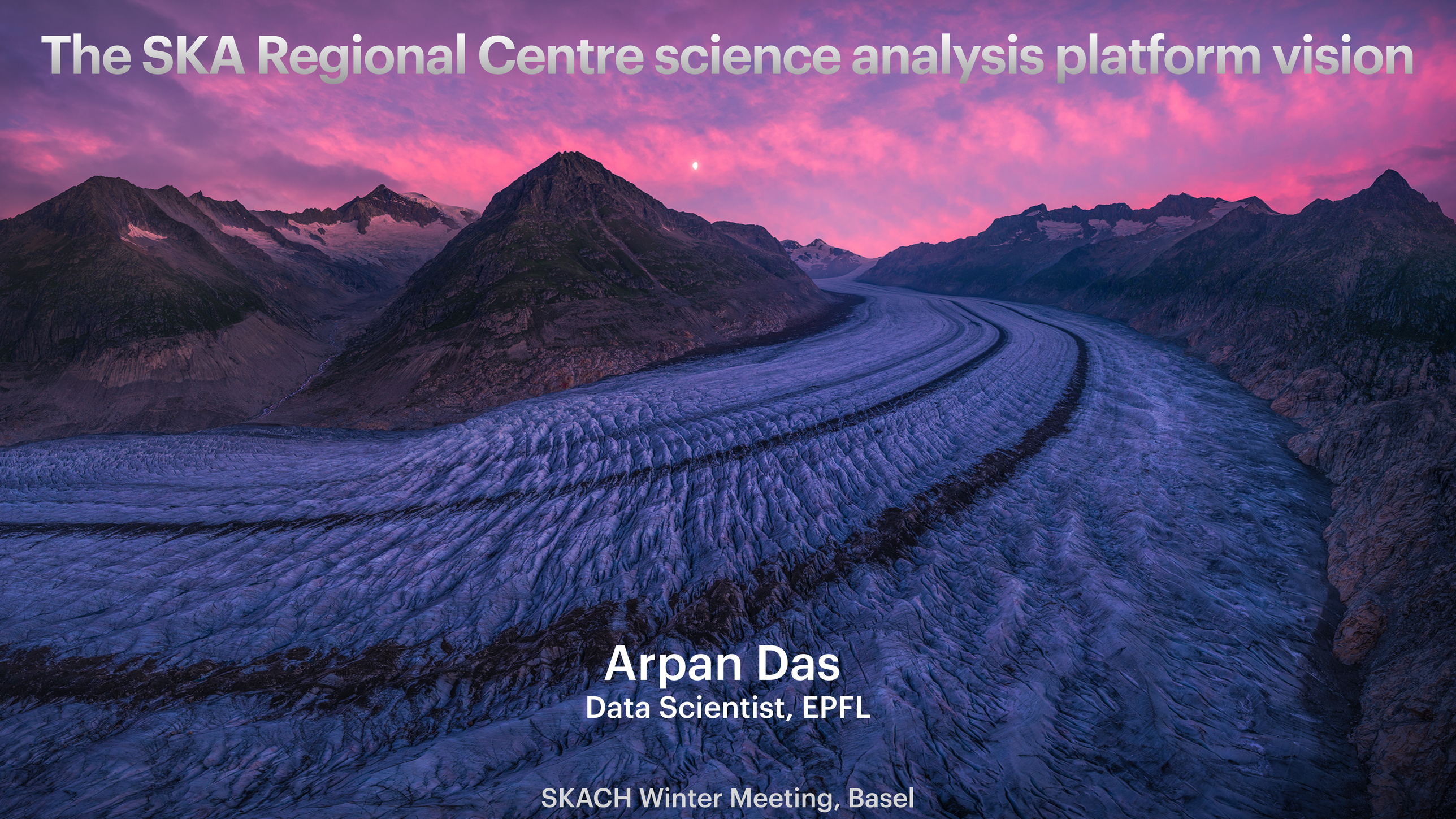
The SKA Regional Centre science analysis platform vision

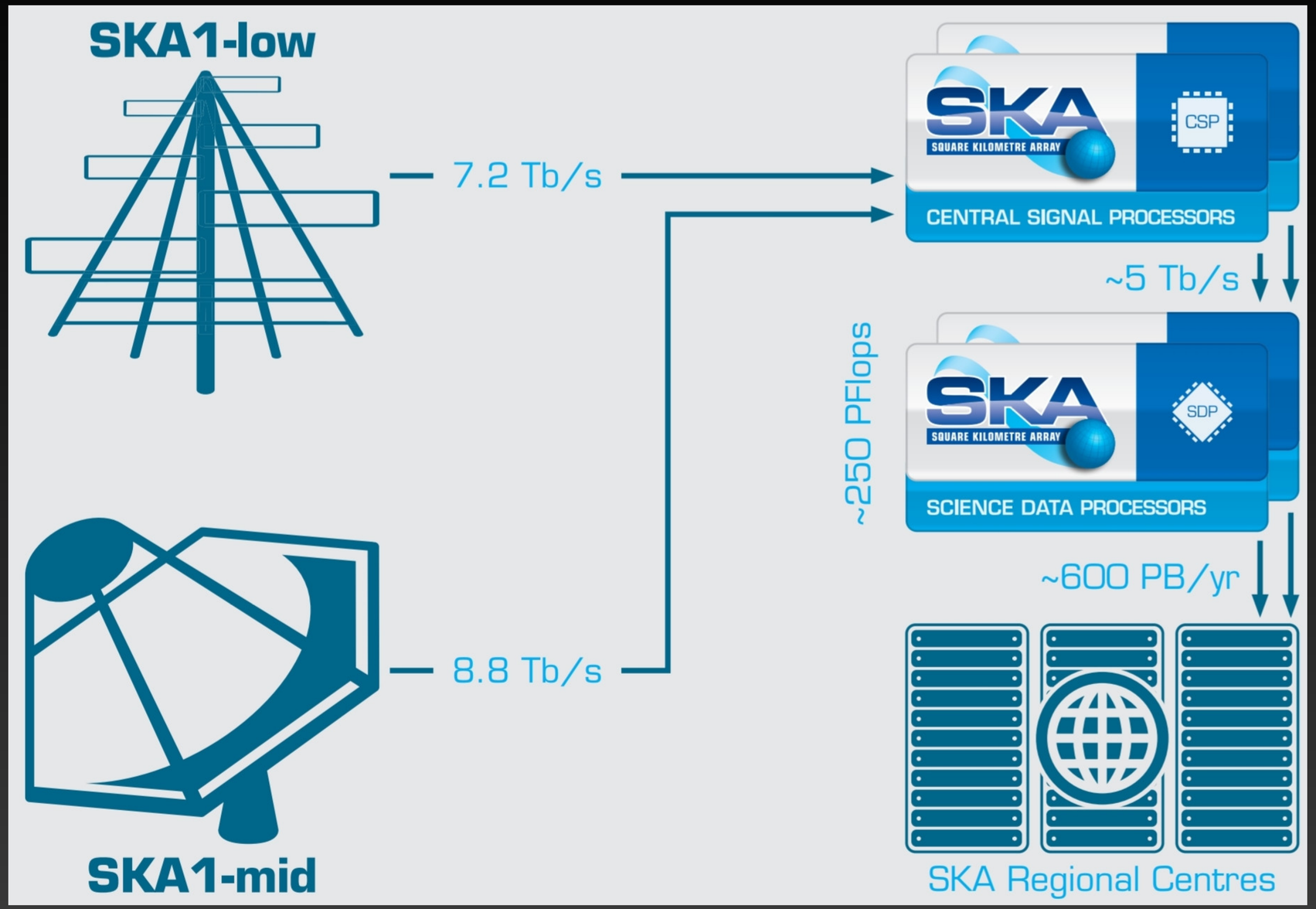


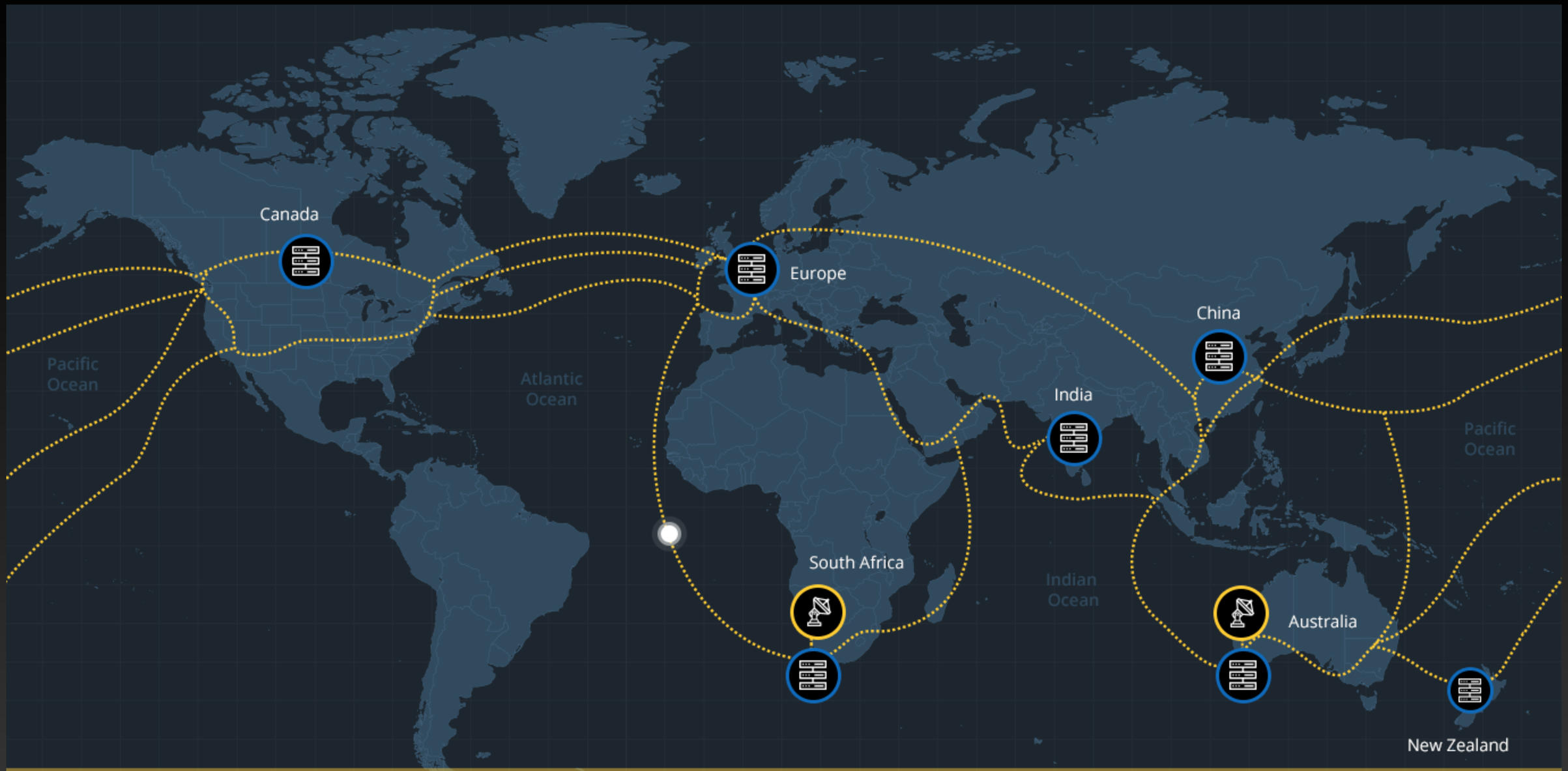
Arpan Das

Data Scientist, EPFL

SKACH Winter Meeting, Basel

SKA DATA


SKA1-low
7.2 Tb/s
SKA
SQUARE KILOMETRE ARRAY
CSP
CENTRAL SIGNAL PROCESSORS
~5 Tb/s
~250 PFlops
SKA
SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSORS
~600 PB/yr
8.8 Tb/s
SKA1-mid
SKA Regional Centres



SKA Regional Centres **(SRCs)**

SKA Regional Centre Capabilities Blueprint

Science Enabling Applications

Analysis Tools, Notebooks,
Workflows execution
Machine Learning, etc

Visualization

Advanced visualizers for SKA data
and data from other observatories

Interoperability

Heterogeneous SKA data from
different SRCs and other observatories

Distributed Data Processing

Computing capabilities provided by
the SRCNet to allow data processing

Data Management

Dissemination of Data to SRCs
and Distributed Data Storage

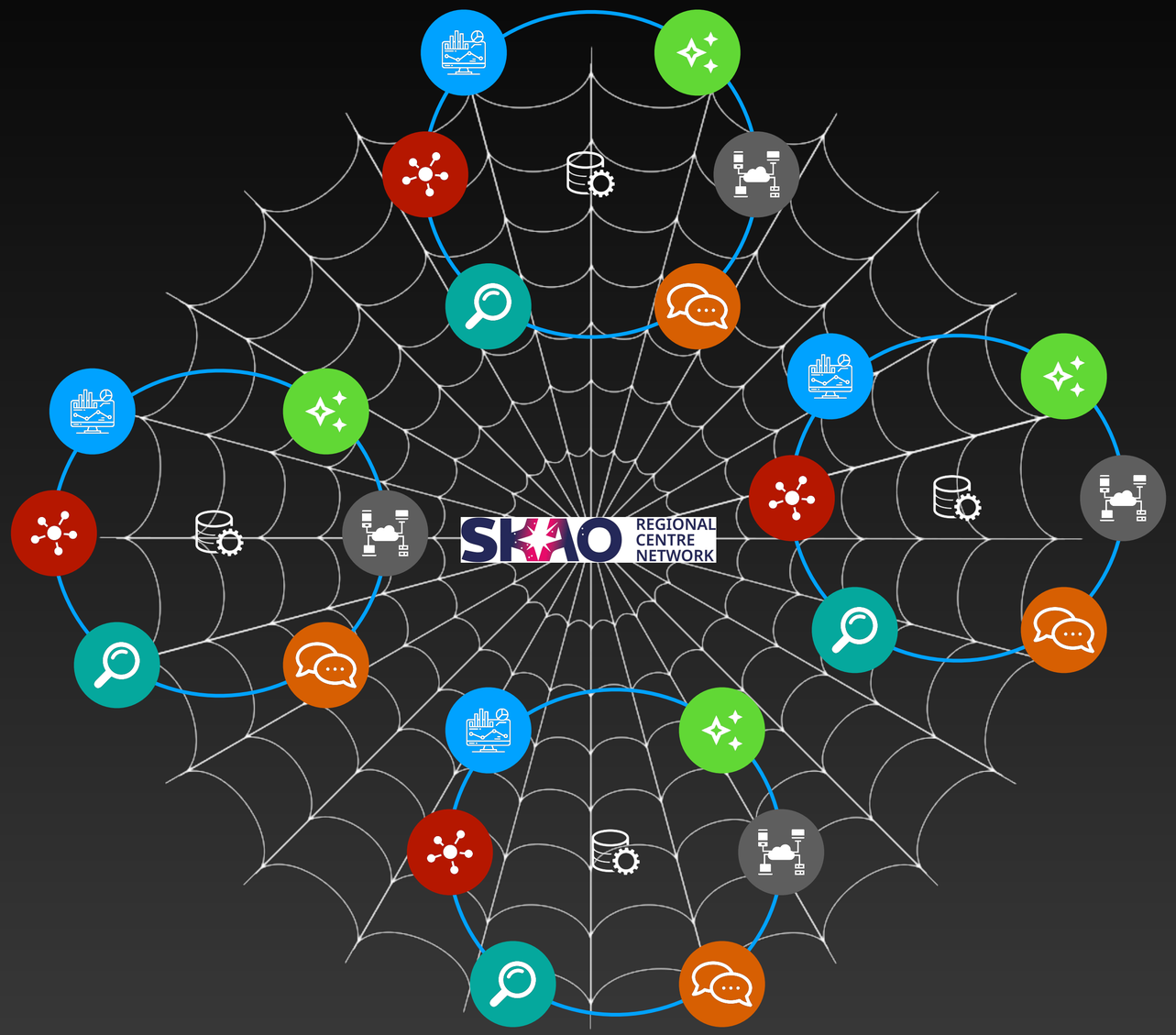
Data Discovery

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

Support to Science Community

Support community on SKA data use,
SRC services use, Training,
Project Impact Dissemination

SRC Network global capabilities



- Every node is an instance of the blueprint
- Interconnections are done using agreed APIs, using FAIR and VO protocols where available
- Collectively meet the needs of the global community of SKA users
- Anticipate heterogeneous SRCs, with different strengths

Team Tangerine

The Tangerine Team is working on science platform components, including prototypes, Data processing Notebooks and Distribution of software, tools and services.

EPFL

University of Victoria

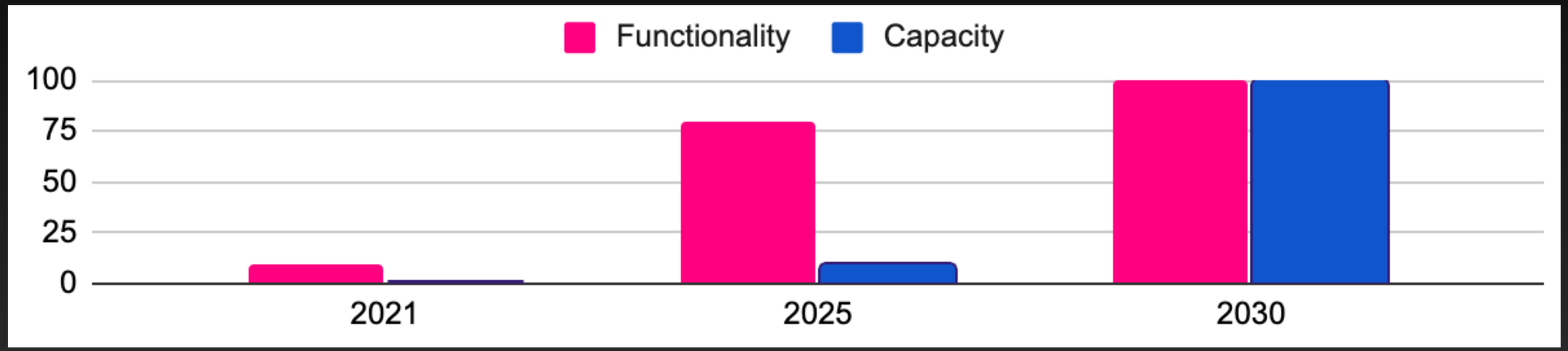
MANCHESTER 1824
The University of Manchester

n|w University of Applied Sciences and Arts Northwestern Switzerland

ASTRON
Netherlands Institute for Radio Astronomy

University of Hertfordshire UH

Timeline



Vision Document Draft Notice

The platform vision is being reviewed by several people within the the SKA SRC network development team. Almost all the internal feedbacks had been already addressed and the draft is currently ready to be distributed among the external reviewers

SKAO REGIONAL CENTRE NETWORK

SRC Science Analysis Platform Vision Document

de Boer, J.[1], Cimpan, I.[2], Das, A.[3], Fabbro, S.[4], Grange, Y. G.[1*], Hardcastle, M. J.[5], Sharma, R.[6], Skipper, C. J.[2], Swinbank, J. D.[1], Webster, B.[5]

[1] *ASTRON, the Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4,7991 PD Dwingeloo, The Netherlands*
[2] *Jodrell Bank Centre for Astrophysics, Alan Turing Building, The University of Manchester, Manchester, M13 9PL, UK*
[3] *École polytechnique fédérale de Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland*
[4] *NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada*
[5] *Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*
[6] *Fachhochschule Nordwestschweiz, Bahnhofstrasse 6, 5210 Windisch, Switzerland*

[*] corresponding author

Abstract. This document describes the vision for the Square Kilometer Array (SKA) Regional Centres Science Analysis Platform. It is intended to set the broad terms of reference for the platform and to provide guidance for both development teams and other stakeholders. Among the features and services that are expected to be included are data querying and discovery tools, some form of notebook interface, user-managed software environments, workflow management, and a comprehensive set of APIs enabling access to all low-level platform functionality. This document is not a design specification, and the features and services described herein will be further refined, or could be discarded, at a later stage of development.

Design Considerations

Attributes:

- consistency
- scalability
- reproducibility
- usability
- reliability

Guiding principles:

- Highly collaborative
- end-to-end
- accessible

ILLUSTRATION: ISTOCKPHOTO

Computing Services

- Analysis through the SRC will be distributed over different regional, national or supra-national compute infrastructures.

- Computing requirements are varied and include distributed processing of very large datasets for generation of ADP or even running of the SKAO's SDP pipelines, as well as user-driven batch processing or interactive processing via, for example, a notebook.

- Resource management will be carried out by SRCnet to make the most efficient use of compute resources. An important feature of the distributed nature of the SKA archive is that for maximum efficiency data/workflows may need to be transferred to the most appropriate node for generating an ADP.

Archive and distributed data

- The SRCNet will be responsible for providing an archive for the released data products, together with the functionality needed to allow users to query the archive, retrieve data for further processing.
- There will be a process for users to upload data products, and their associated metadata, to the archive, where they can be made visible to other users.
- Once data is submitted to the archive, it will not be possible for users to alter the archival contents, although it will be possible to upload a new version of the data to the archive
- The users will be able to run queries on the archive using appropriate query languages, e.g. TAP, ADQL

Main User Interfaces

- The main user interface for the platform will be a web page (the `portal') providing access to the functionality of the SRC node through a range of services.

- The user can sign on to the portal on the front page using single sign-on criteria and will then be given access to the full user interface; without signing on there will only be limited public data access to, for example, image previews and catalogs

- The user interface will be consistent across different SRC nodes, although in limited circumstances SRC nodes may make minor modifications to the portal to make local resources available.

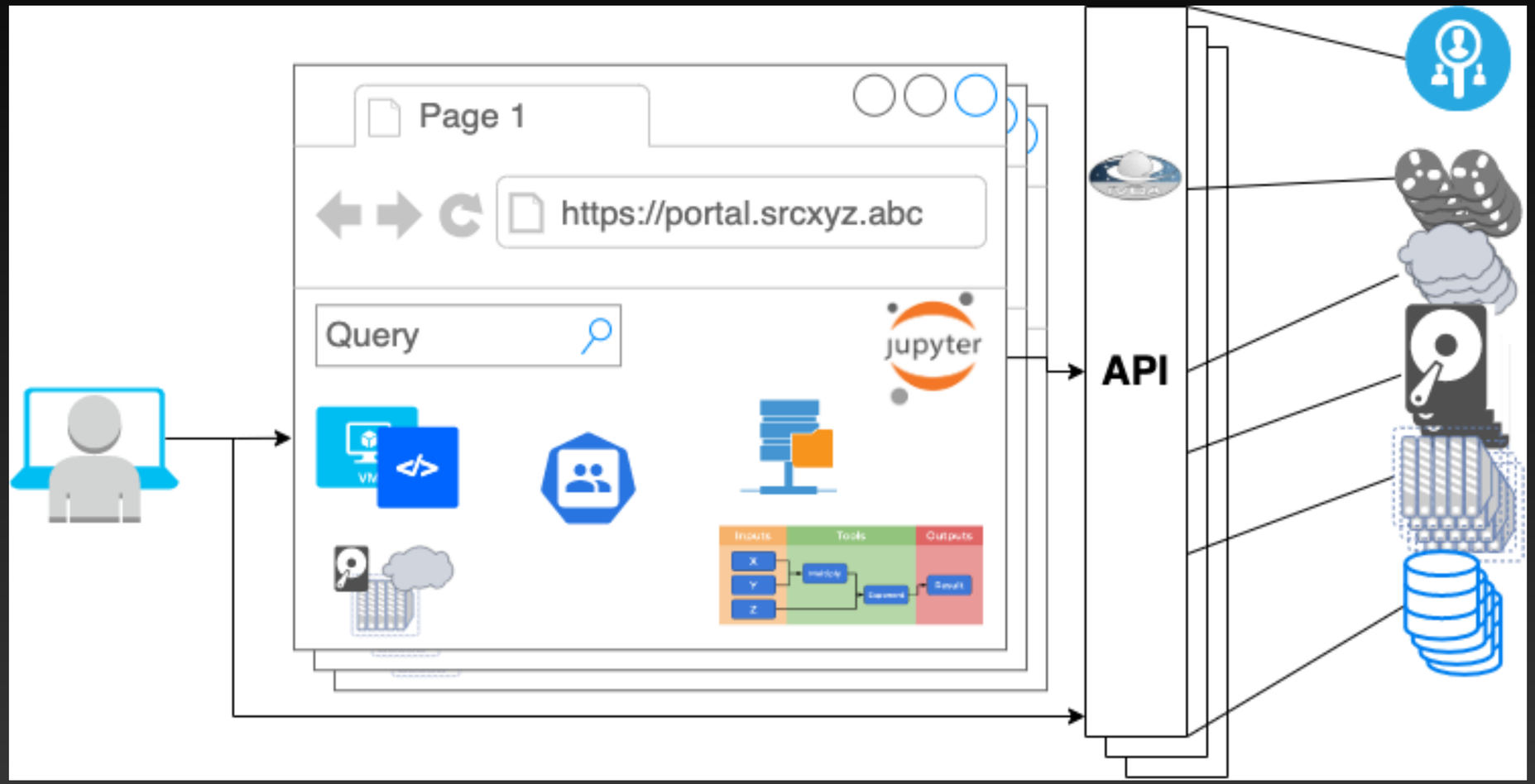
Different Mechanism

- Web-based search engine: The archive search will allow users to search both the ODP and publicly available ADP as well as compatible non-SKA archives.

- Drill-down browsing: The user will be presented with a list of data collections, and can explore these collections by clicking through to a lower level.

- Graphical selection: The user will be shown a cutout of a region of sky, and can select data by clicking on regions of this cutout. This process for data discovery is similar to that used by the Virtual Observatory.

Functionality



The platform in the SRCNet

- The platform brings together all components of the SRC network for the users