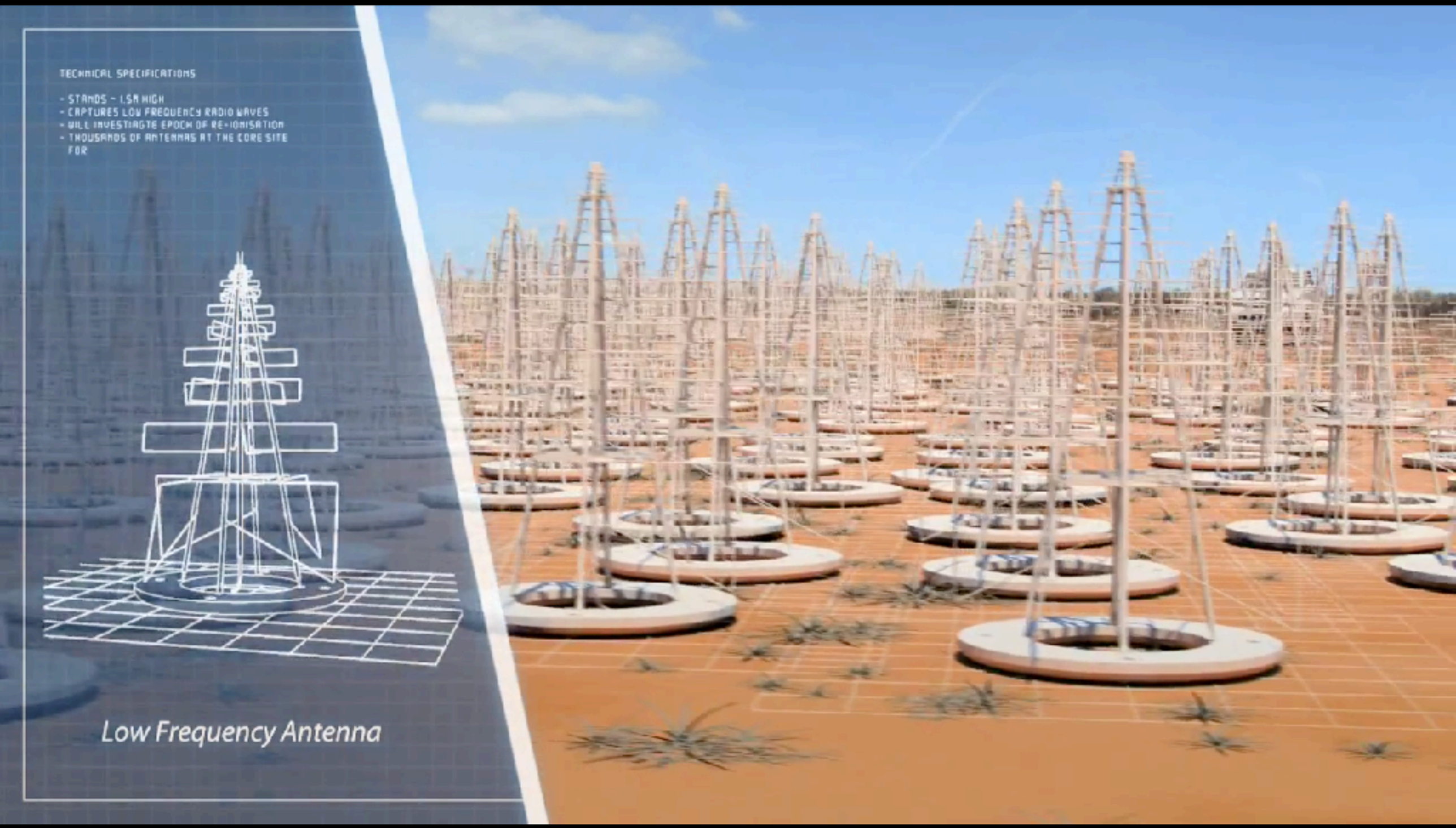
SKA & AENEAS

Anna Scaife
Jodrell Bank Centre for Astrophysics
University of Manchester


TECHNICAL SPECIFICATIONS
- STANDS ~ 15M HIGH
- CAPTURES LOW FREQUENCY RADIO WAVES
- WILL INVESTIAGTE EPOCH OF RE-IONISATION
- THOUSANDS OF ANTENNAS AT THE CORE SITE FOR
Low Frequency Antenna

Future SKA Science Archive

2017
2023

searches on Google 98PB

uploads to facebook 180PB

CERN 73PB

LOFAR Long Term Archive 25PB

YouTube 15PB

6PB

NASDAQ 3PB

PER YEAR
- 1 Petabyte

SKA
Phase1 Science Archive
300PB

Where’s your
latest milestone?

Where’s your latest milestone?

Aeneas

Design and specification of a distributed, European SKA Regional Centre to support the pan-European astronomical community in achieving the scientific goals of the SKA

Advanced European Network of E-infrastructures for Astronomy with the SKA

EC Horizon 2020 (€3 million)

13 countries, 28 partners, SKAO, host countries, e-infrastructures (EGI, GÉANT, RDA), NREN's

Three year project (2017-2019)

- WP1: Project Management
- WP2: Governance Structure and Business Models
- WP3: Computing and Processing Requirements
- WP4: Data Transport and Optimal European Storage Topologies
- WP5: Data Access and Knowledge Creation
- WP6: User Services

AENEAS All-hands Meeting 2018, Observatoire de Cote d'Azure

OBJECTIVES

- Develop a set of **design recommendations** for the ESDC pertinent to (1) data handling strategy, (2) scientific functionality and (3) software environment.
- Produce a **high level architectural design** for the ESDC with a **sizing** and **costing** estimate.
- Provide **supporting verification work**, including both theoretical analyses and direct prototyping of critical elements.
- Identify **gaps**, highlight **risks** and make recommendations with respect to **mitigation**.

WORK BREAKDOWN

T3.1 ESDC Processing: Inventory of SKA science cases and post-SDP computing requirements

T3.2 ESDC Data storage: Inventory and sizing of SKA science data products and ESDC user-derived products

T3.3 Evaluation of existing HPC, cloud and distributed computing technologies

T3.4 Design and costing for distributed ESDC computing architecture

T3.5 Requirements for interfaces to SKA Science Archives & Other Repositories

T3.6 Validation, Verification & Proof of concept activities utilizing SKA pathfinder and pre-cursor facilities

SKA Observations

Scientific Value

Multiple Use Cases

The SKA Project has:

11 Science Working Groups

Extragalactic Spectral Line
Our Galaxy
Solar, Heliospheric & Ionospheric Physics
Epoch of Reionization
Cosmology
Extragalactic Continuum (galaxies/AGN, galaxy clusters)
Cradle of Life
HI galaxy science
Magnetism
Pulsars
Transients

2 Science Focus Groups

High Energy Cosmic Particles
VLBI

Scientific Use Case —> Pilot Compute Models

- Implementing example use cases on existing infrastructure
- Mostly using WLCG resources
- Identifying potential bottlenecks in e-infrastructure designed for other fields

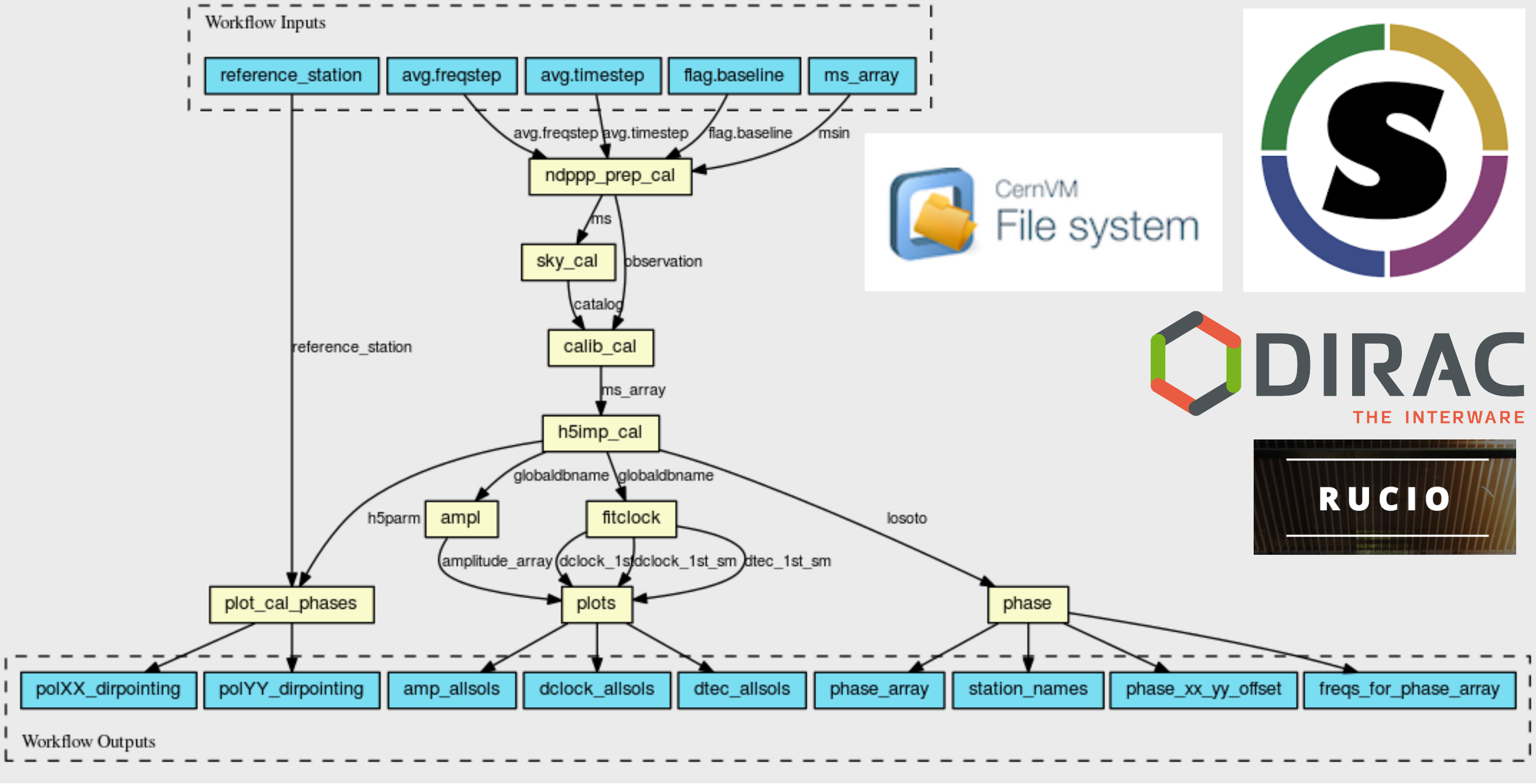
Current Compute Model Use Cases:

- Calibration & Imaging Use Case
- Image-based Object Detection & Classification Use Case
- Catalogue-based Cross Matching incorporating External Archives Use Case

- Image Mosaicking Use Case
- Image Cube Stacking Use Case
- Time-domain Re-folding Use Case

skatelescope.eu VO

Use Case : Calibration & Imaging

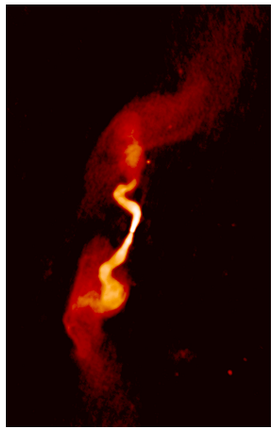


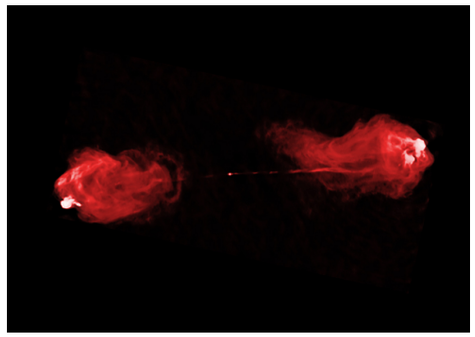
CernVM File system

DIRAC THE INTERWARE

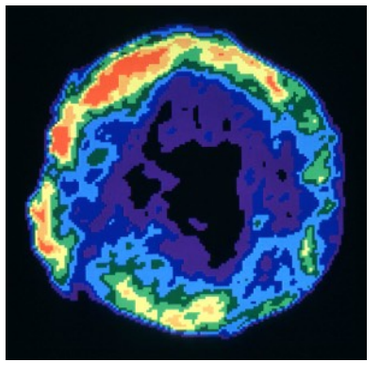
RUCIO

Radio Galaxies



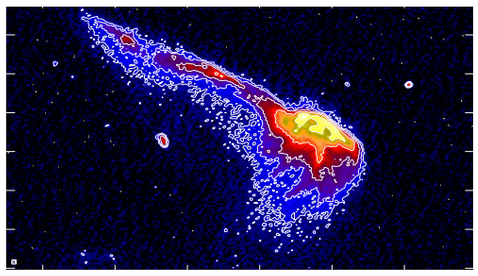


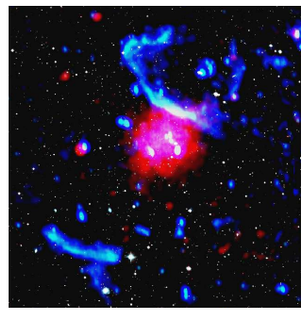
Supernova Remnants



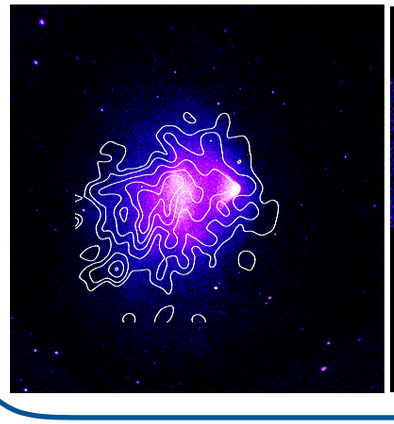
Many Source Classes

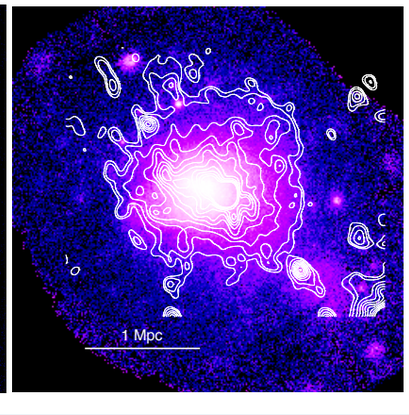
Radio Relics



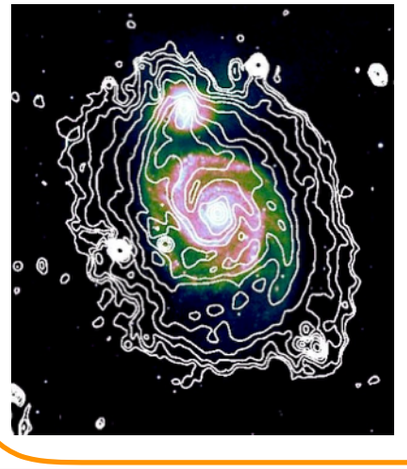


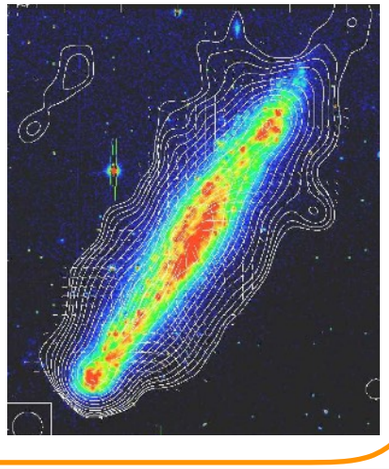
Radio Halos

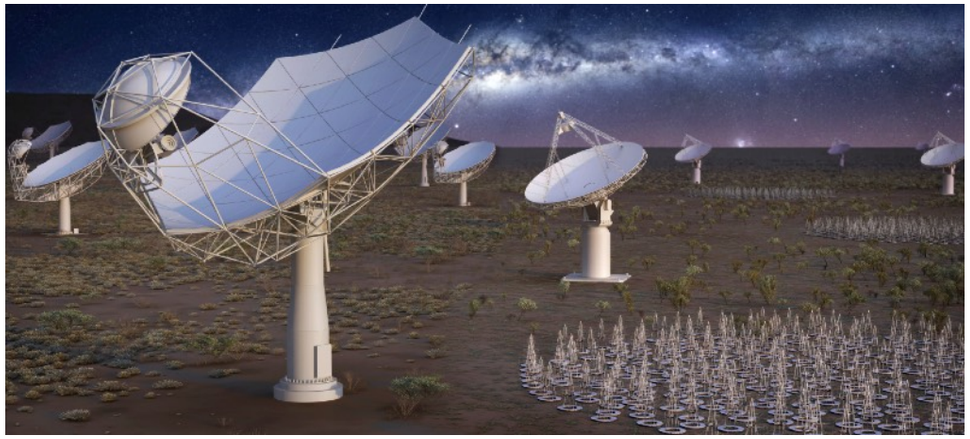




Galaxies



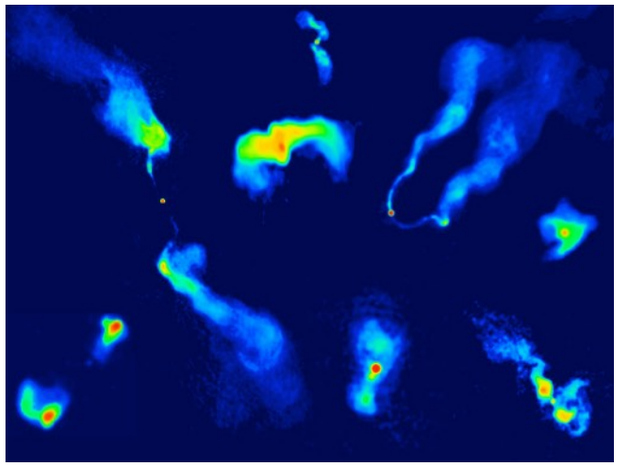


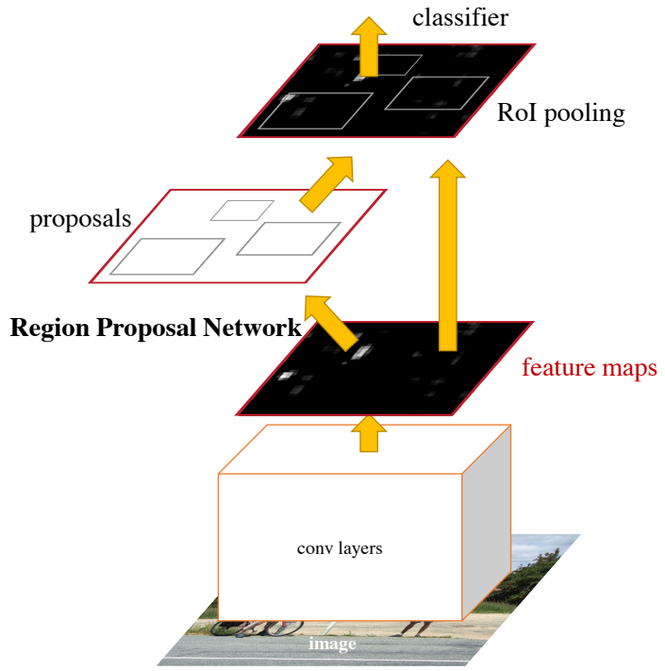


- **Radio galaxy zoo: It took 1 year for ~7000 people to classify 53 229 images**
- **SKA surveys expected to detect >10^7 AGN**
- **Classifying sources by eye takes too much time!**

Banfield et al 2015

Use Case : Object Detection & Classification

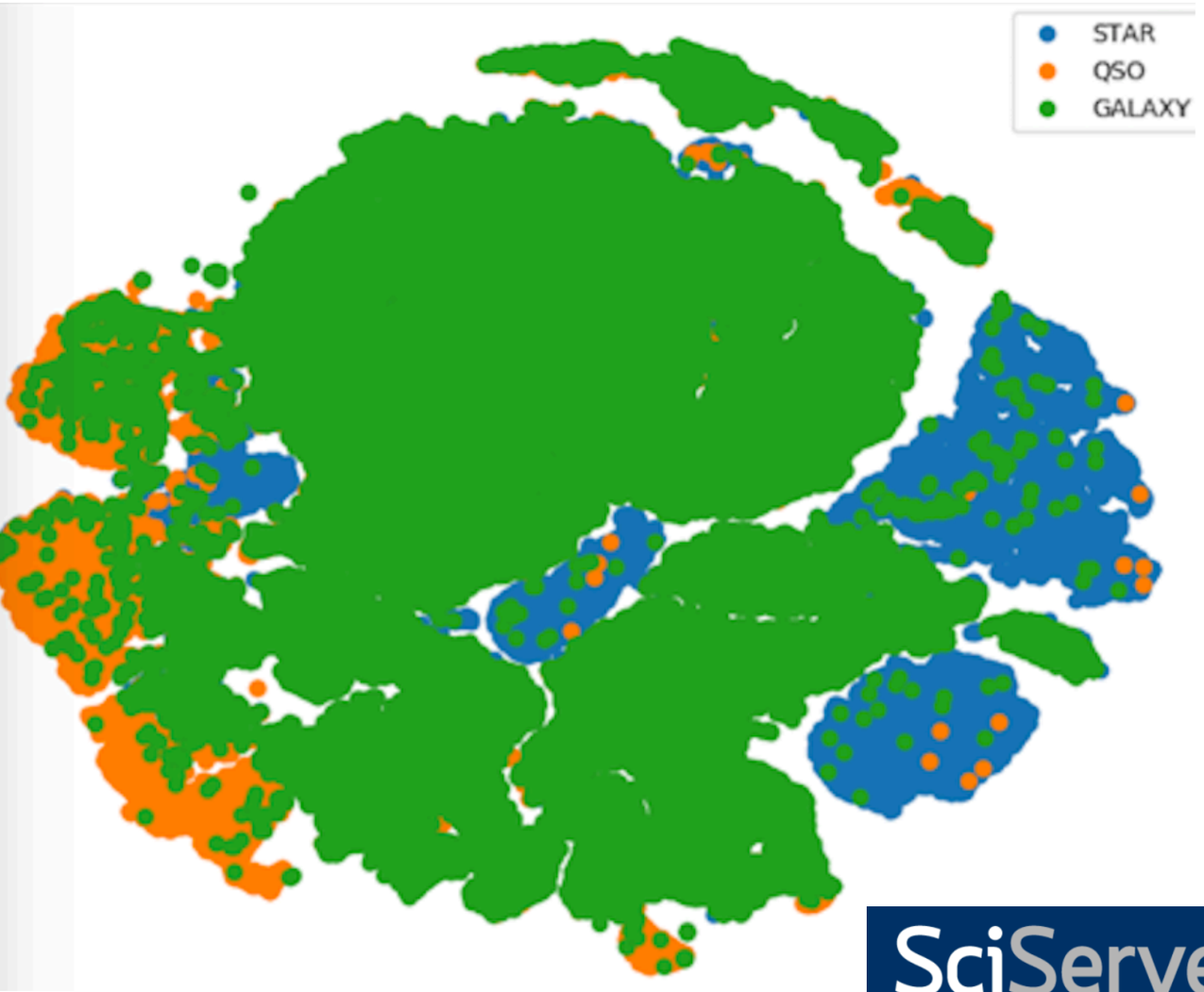




Automatically generate source catalog including source classification


WHAT IS
THIS?

Use Case : Classification using External Archives



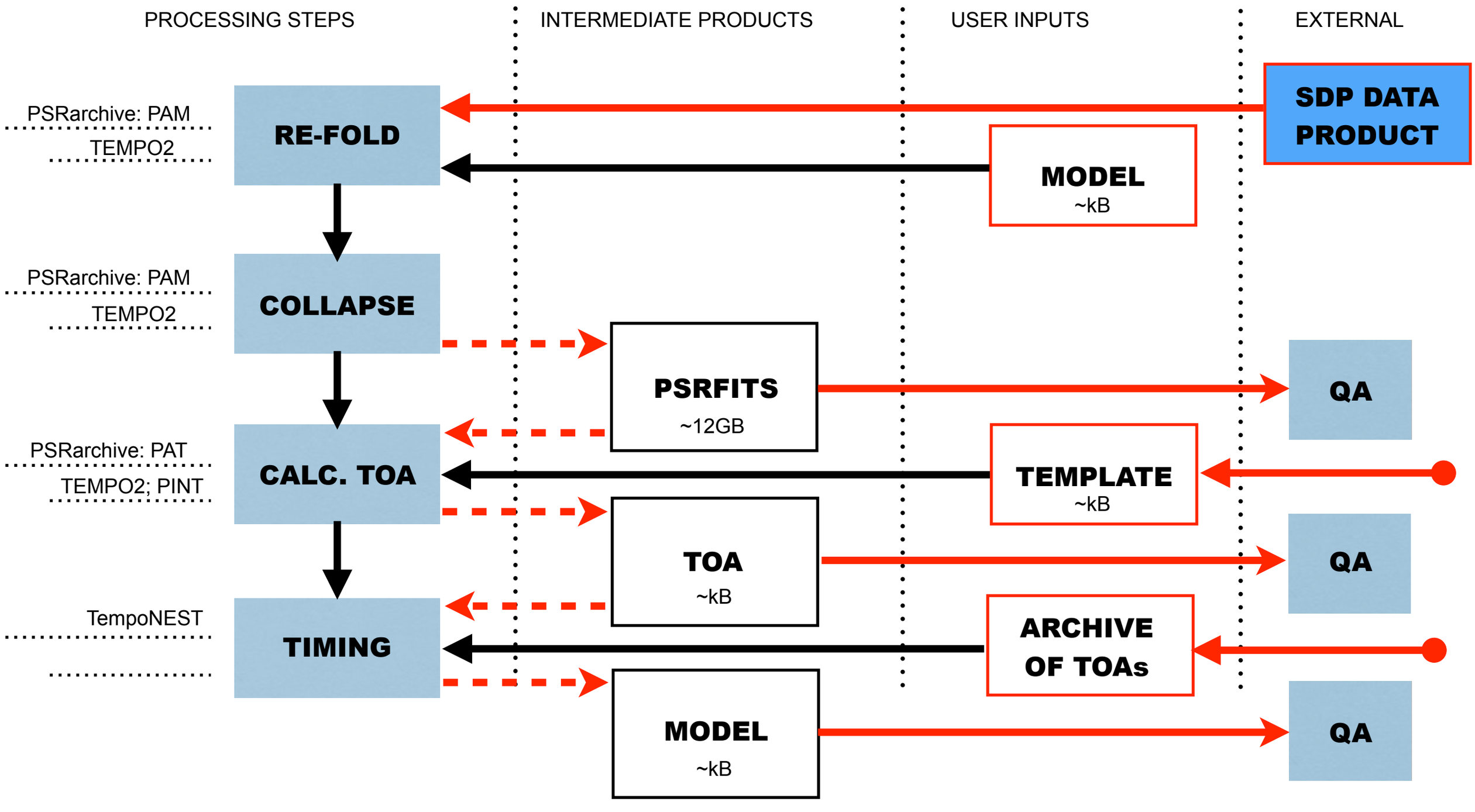
- **2.5 million sources**
- **Provide photometry data (optical/IR/radio)**
- **Similar sources cluster, learning source types**
- **Afterwards colour-coding the plot by type shows galaxies/stars/quasars**
- **97 % accuracy on classifying stars/galaxies/QSO**

SciServer

SkyServer

SkyServer provides browser-based public access to all Sloan Digital Sky Survey data for everyone from teachers to professional astronomers.

COMPUTE MODEL USE CASE : Pulsar SWG



Aeneas