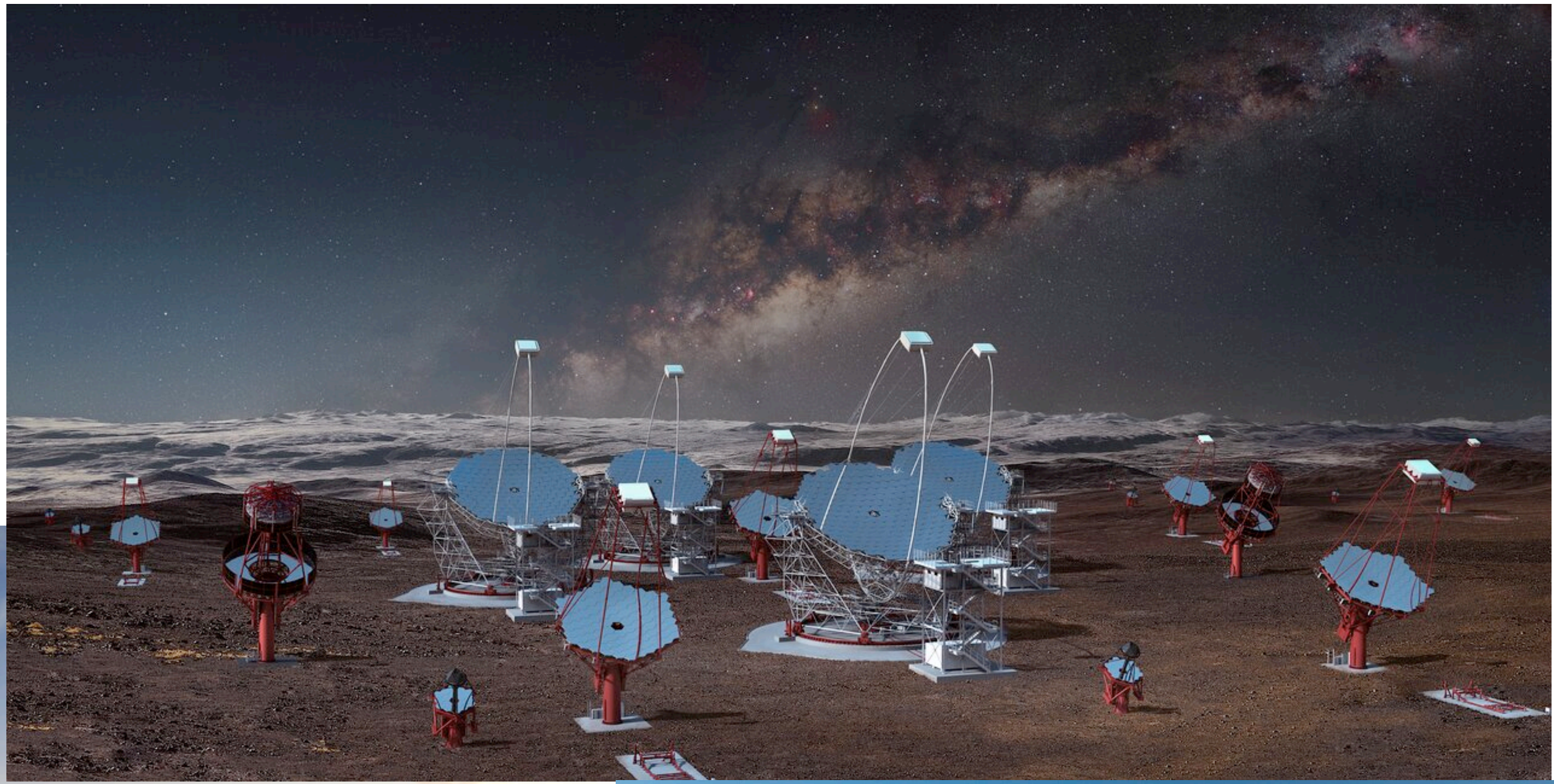
SKA/CTA synergy on scientific computing



Jean-Paul Kneib (EPFL)

EPFL

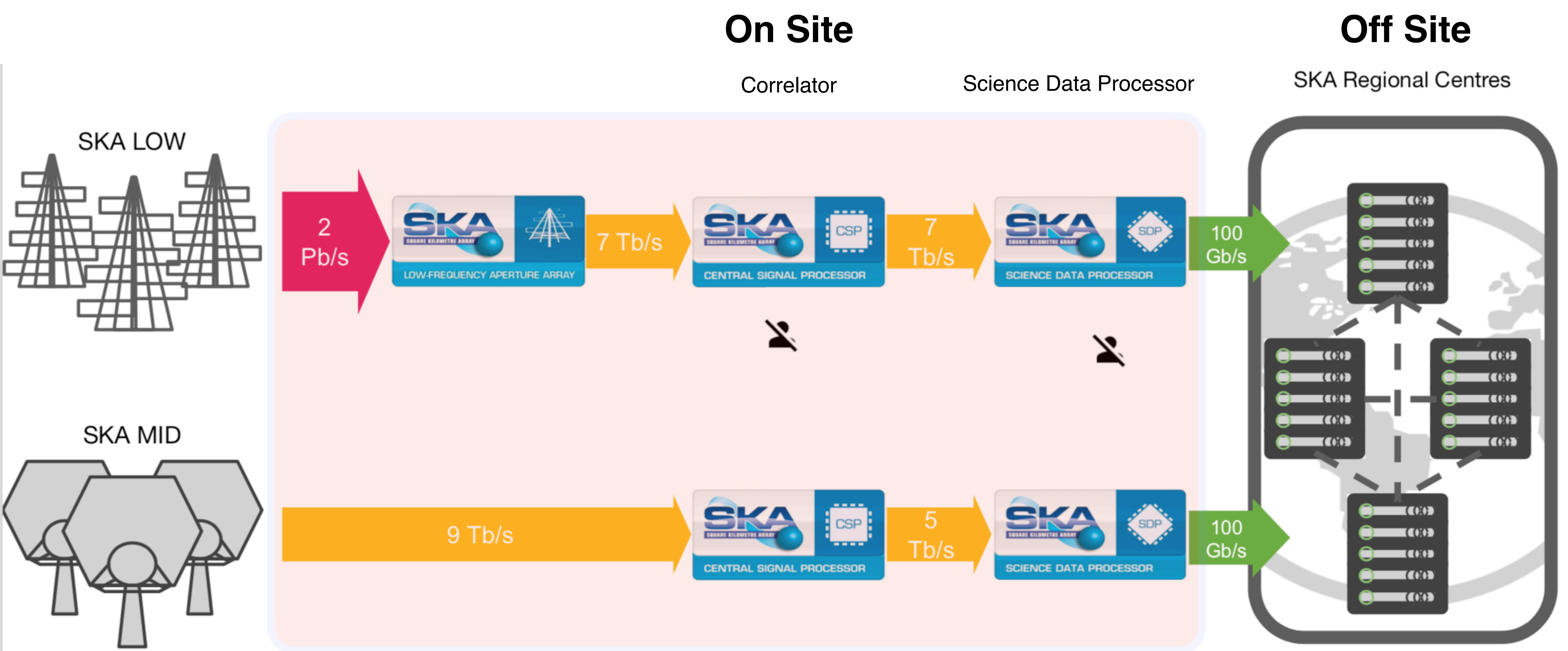
Outline

- SKAO status
- Synergies
  - Sky Simulations
  - HPC pipelines for massive dataset
  - Online Data Analysis Tool
  - Science: e.g. Transients Analysis

SKAO status

- Feb 2021: the Square Kilometre Array Observatory (as an Inter-Governmental Organisation) started
- June 2021: EPFL as the leading house for Switzerland joined SKAO
- July 2021: Approval of SKAO construction
- Dec 2021: The Swiss Government signed the contract to join SKAO
- 2022: Switzerland is a full member of the SKAO
- ~2025: First SKAO dataset

SKAO Data Flow



From Miles Deegan (SKAO)

SDP - data flow, architecture, functions

Ingest ~ 1 TB/s - all the time

~130 PFlops (peak)

Buffer ~50 PB, mix of ‘hot’ and ‘cold’. Need to read at ~10 TB/s

Dedicated 100 Gb/s link to outside world

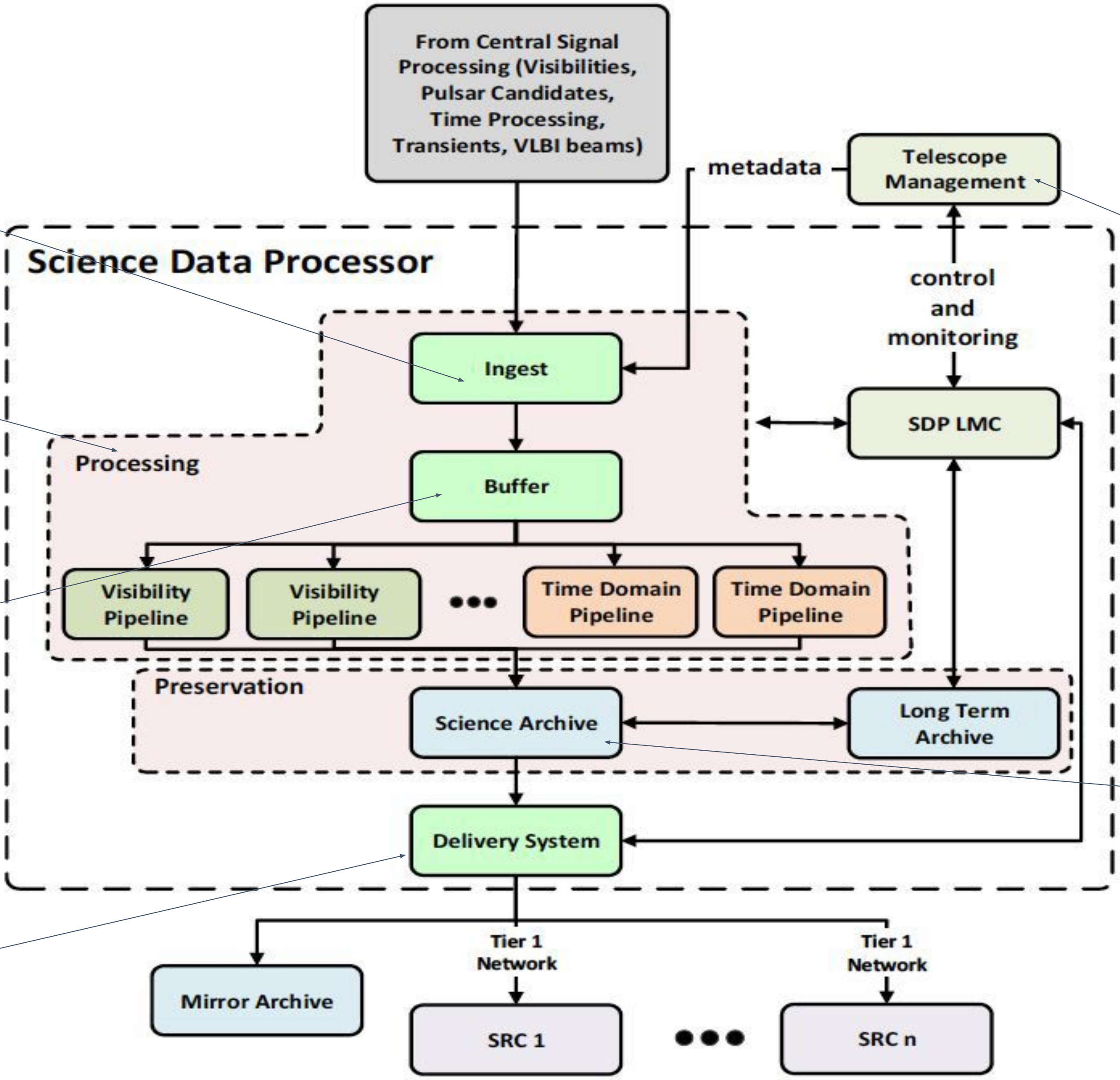


Figure 1-16: SDP data-flow, architecture and functions.

System ultimately under the control of TMC

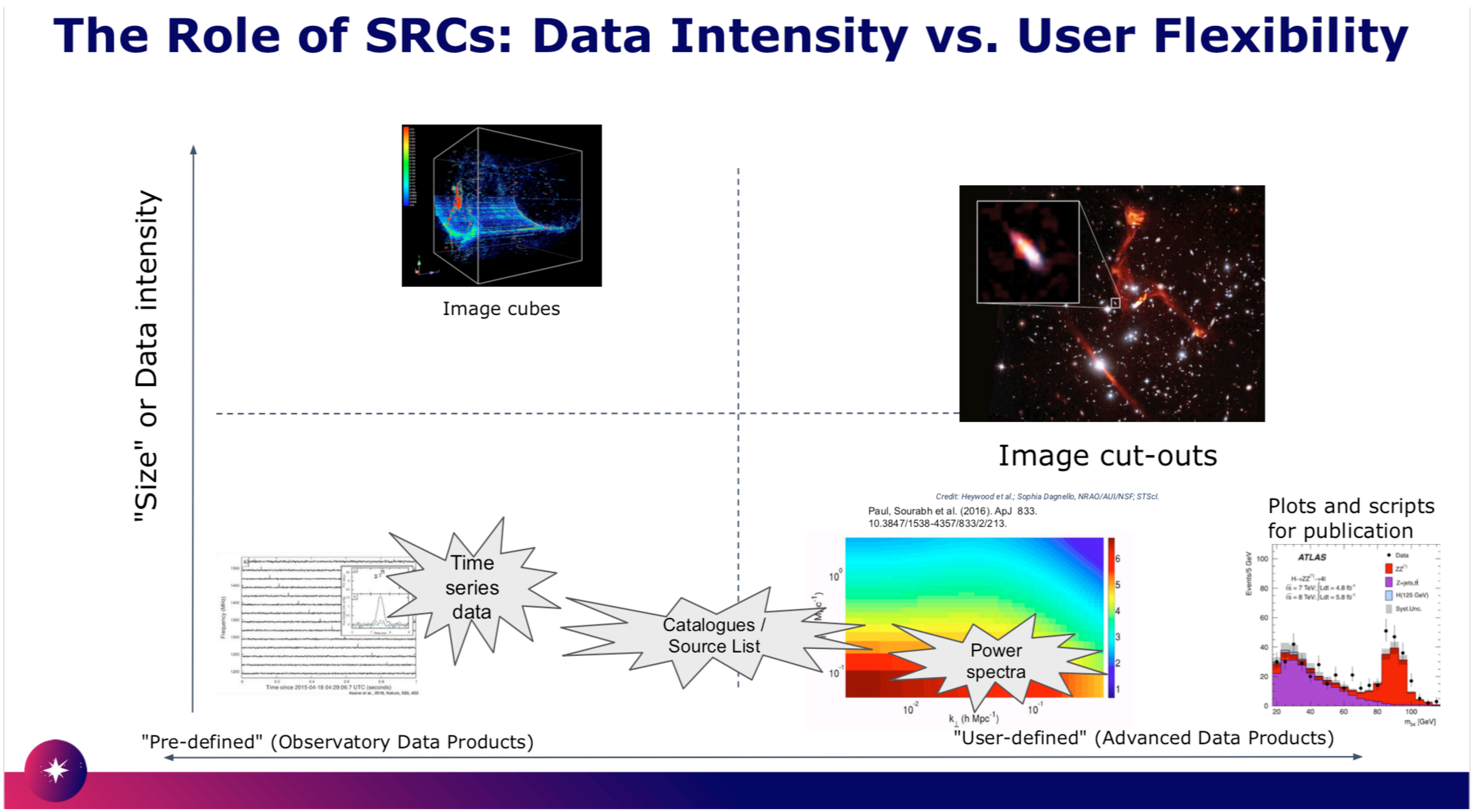
~300 PB/year/SDP in full operations

SKAO Regional Center (SRC)

From Rosie Bolton (SKAO)

SRCs will provide **collaborative online tools** (workflows, notebooks) backed up by powerful compute and data management.

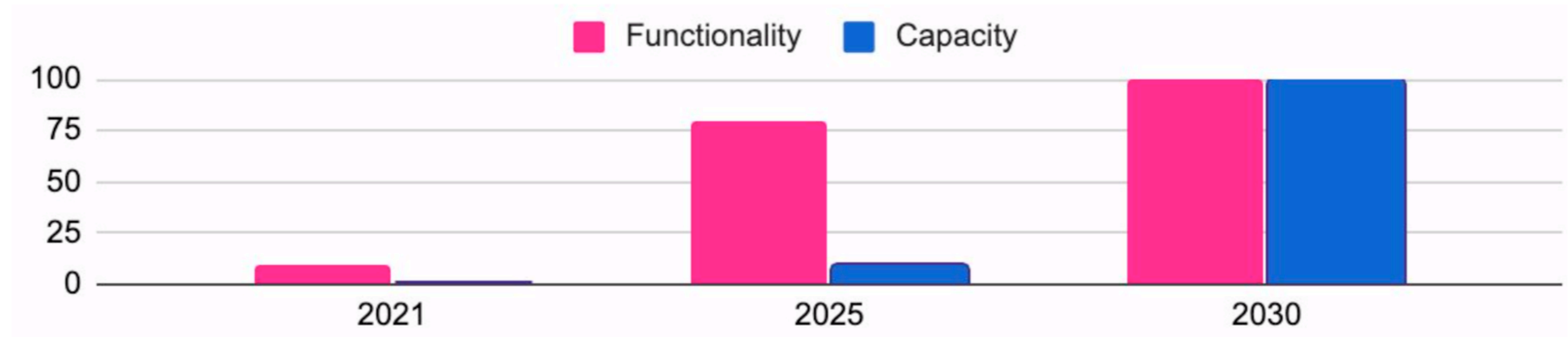
SKA data will eventually become public => ***largest data science repository* - up to 700 PB/year**



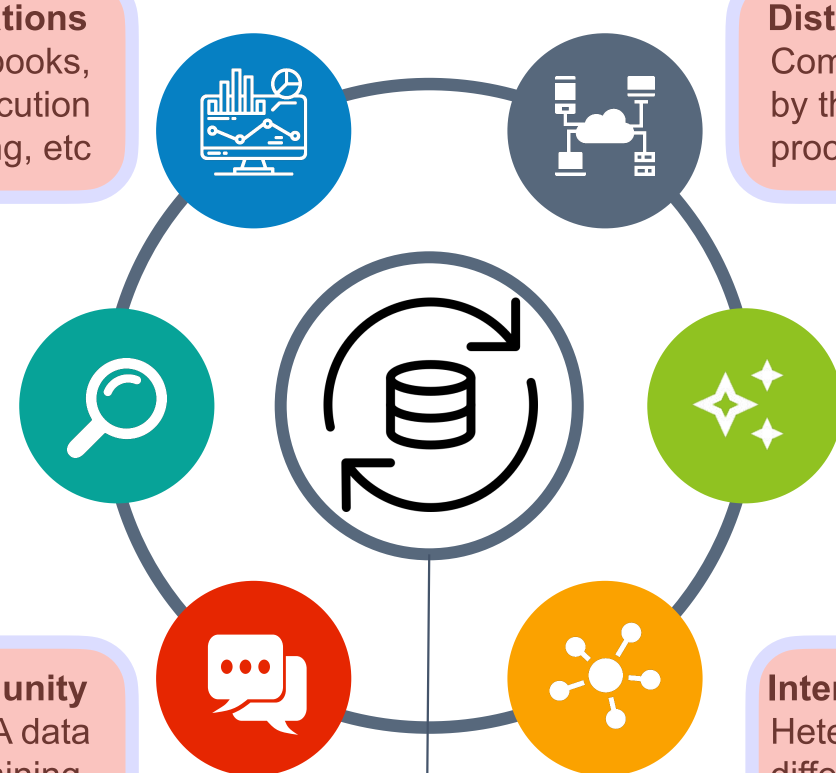
SKAO Regional Center (SRC)

From Rosie Bolton (SKAO)

Timeline



SKA Regional Centre Capabilities



Science Enabling Applications
Analysis Tools, Notebooks, Workflows execution Machine Learning, etc

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 visualizers for SKA data and data from other observatories

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

Interoperability
Heterogeneous SKA data from different SRCs and other observatories

Data Management
Dissemination of Data to SRCs and Distributed Data Storage

SKAO/CTAO computing synergies

- **Common challenges:**
  - PBytes of data to handle every year
  - Current algorithm that do not scale (yet)
  - Need of large numerical/Monte-Carlo simulations to analyse and interpret the data
  - *Scientific goals:* e.g. transient phenomena (AGN, GRB, Gravitational Waves, FRB/magnetars? …), but also cosmology.

Sky simulations for SKA

From Lucio Mayer (UZH)

Cosmological <u>hydrodynamical simulations</u>
build virtual Universe datasets modelling cold neutral gaz

<u>Exascale codes</u> for next generation cosmological hydrodynamical simulations: SPH-EXA
Co-funding by Platform for Advanced Scientific Computing (PASC)

Synthetic observations of simulated datasets <u>aided and accelerated by machine learning</u>

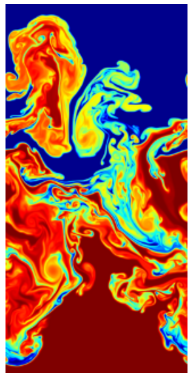
=> Sky Simulation could also be adapted to match CTA data interpretation needs:

* Constrained simulations - matching local observations
* Magneto-HydroDynamical simulation to properly account of the B-field.

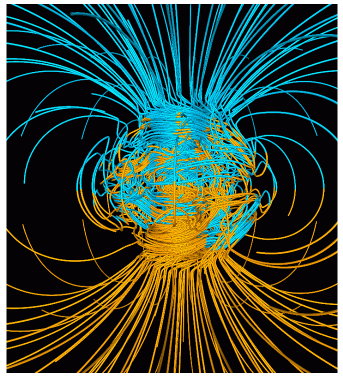
SPH-EXA: Smoothed Particle Hydrodynamics at Exascale

From Lucio Mayer (UZH)

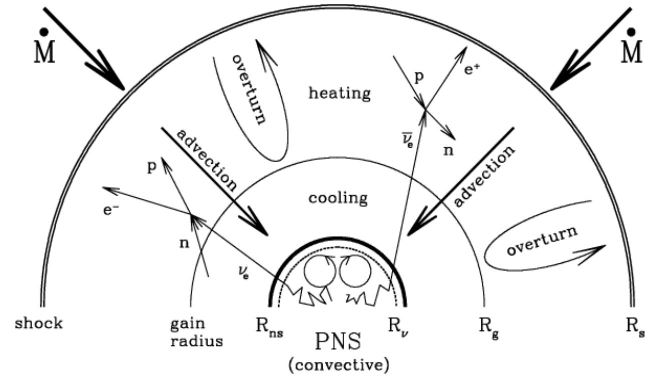
University of Basel | University of Zurich | PASC Platform for Advanced Scientific Computing | CSCS Centro Svizzero di Calcolo Scientifico Swiss National Supercomputing Centre



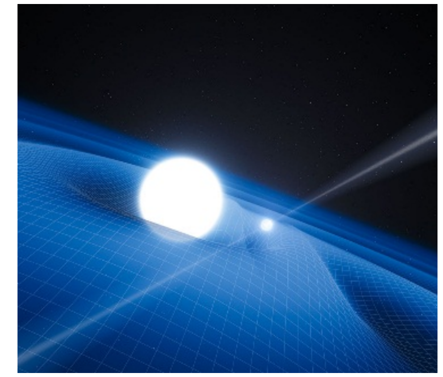
Hydrodynamics



Magnetic fields

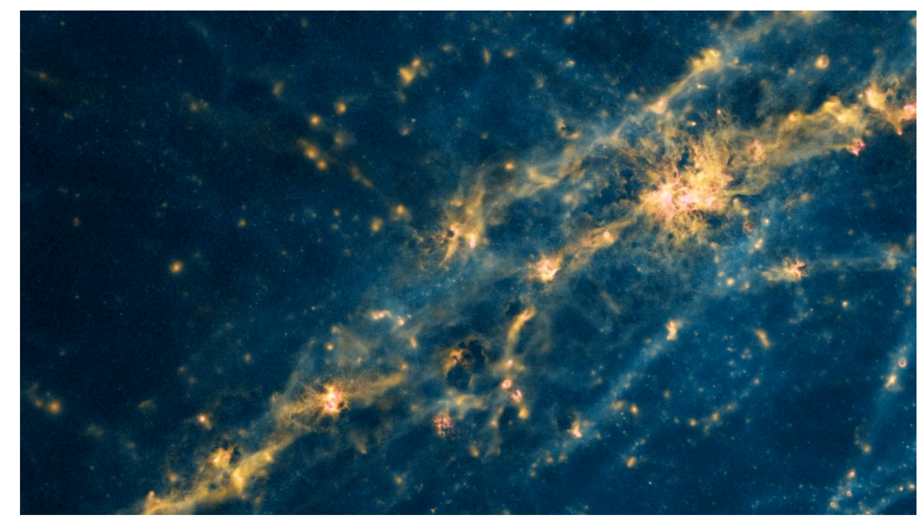


Neutrino transport



General Relativity

Nuclear Physics



Galaxy formation

Vision

Development of a scalable and fault tolerant SPH framework that executes at Exascale to perform for the first time trillion particle galaxy formation simulations with SPH, gravity, and radiation (ExaPHOEBOS).

SPH-EXA: From Mini-App to Modular Production Code

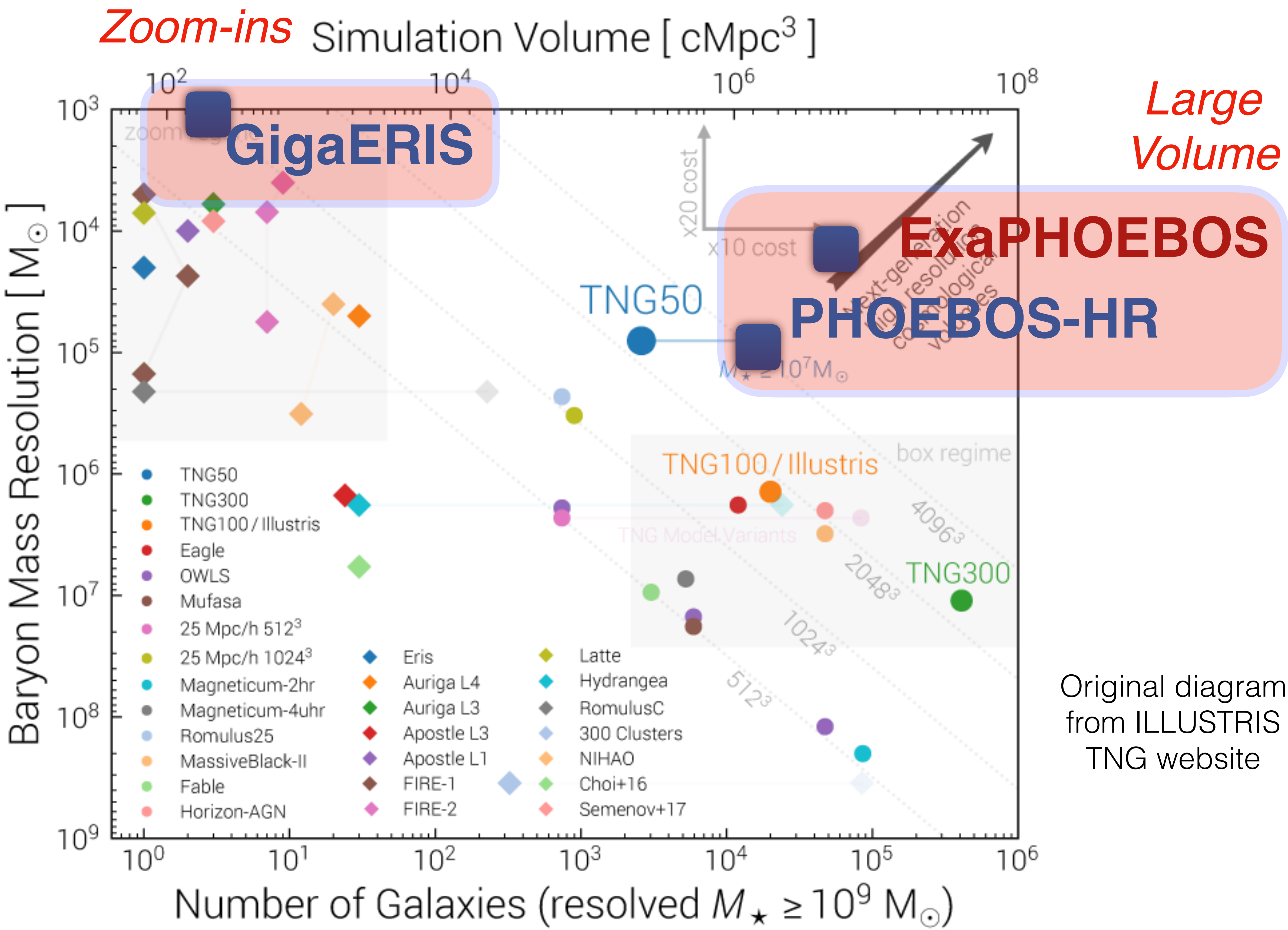
Synthesizes the characteristics of state-of-the-art SPH codes
Provides an MPI+X (OpenMP/OpenACC/CUDA) reference optimized C++ (header only) implementation
Employs adaptive load balancing and fault-tolerance
Implements basic to advanced SPH operands
Explores new programming paradigms (e.g., HPX)
Works with Cray, Clang, GCC and Intel compilers
70% efficiency at 65bn particles on 2048 GPU nodes @ Piz Daint

SPH-EXA project:
Project (2017-2020) + extension 2021
Continues as SPH-EXA2 (2021-2024)
Combines with the Swiss participation in the SKA Observatory (2021-2025)
Targets simulation > 1 trillion particles with SPH, gravity, radiation

https://hpc.dmi.unibas.ch/en/research/pasc-sph-exa2/
https://hpc.dmi.unibas.ch/en/research/sph-exa/

New generation cosmological hydro simulations (PizDaint and Alps/Eiger)

From Lucio Mayer (UZH)



Original diagram from ILLUSTRIS TNG website

HPC pipelines on Massive DataSet

- Current algorithms that do not scale !
  - Adapt or redesign codes to parallel computing on HPC benefiting of both CPUs and GPUs (CSCS
  - Use Machine Learning Techniques (heavily relaying on GPUs) for *simulations or source identifications* (e.g. Gamma vs. Cosmic Rays in CTA images)
  - Benchmarking new codes/developments

Computational Imagine of interferometric data

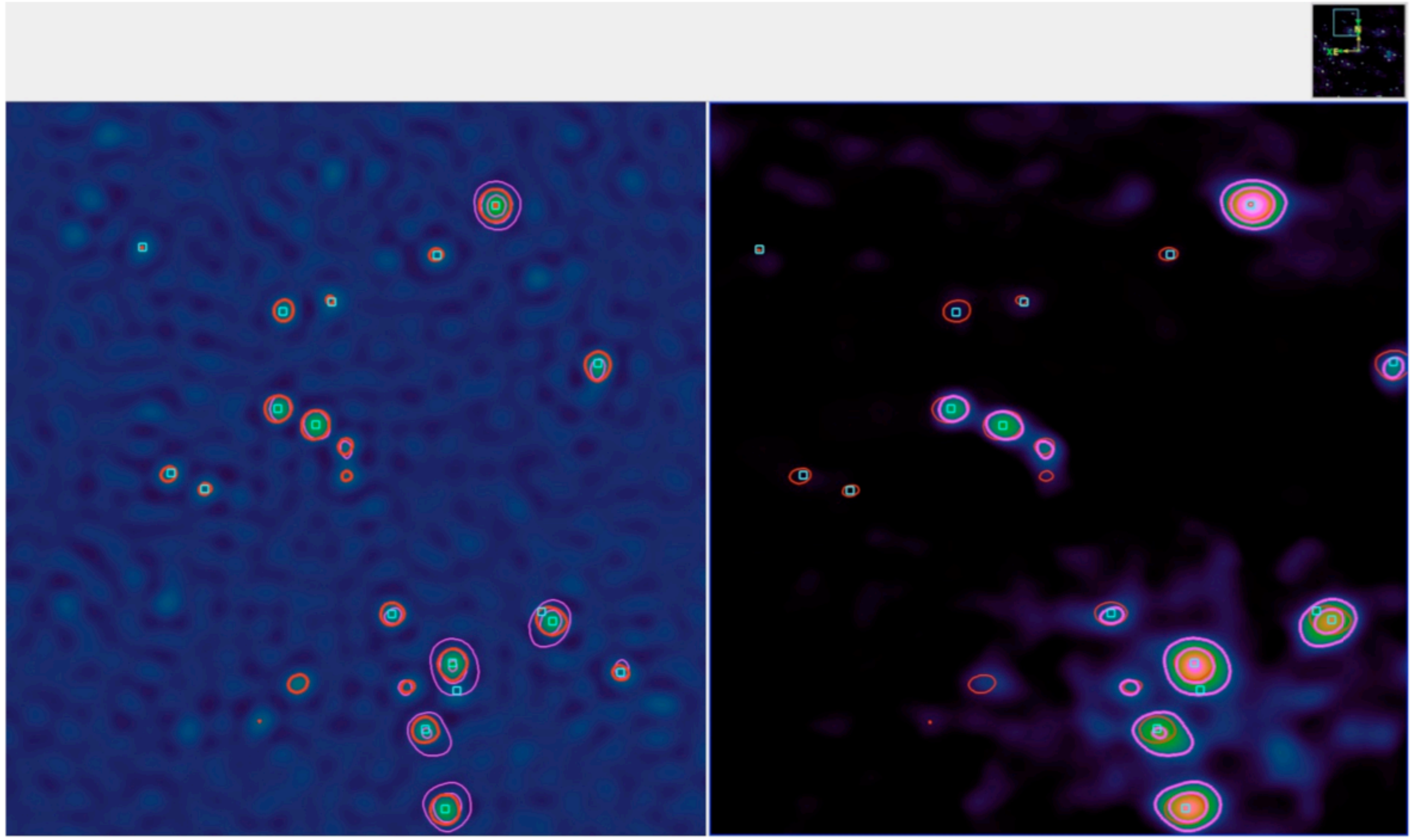
From Matthieu Simeoni (EPFL)

Developing: fast computational imaging codes that are more scalable than standard codes.

Algorithmic: sketching in time for more robust fPCA, better clustering of eigenlevels, BIC-based thresholding for denoising, BB + beamforming.

Acceleration: leverage Hermitian symmetry in NUFFT, faceted NUFFT, GPU implementation.

Validation: comparison to CLEAN, accuracy of NUFFT, fPCA on extended sources?

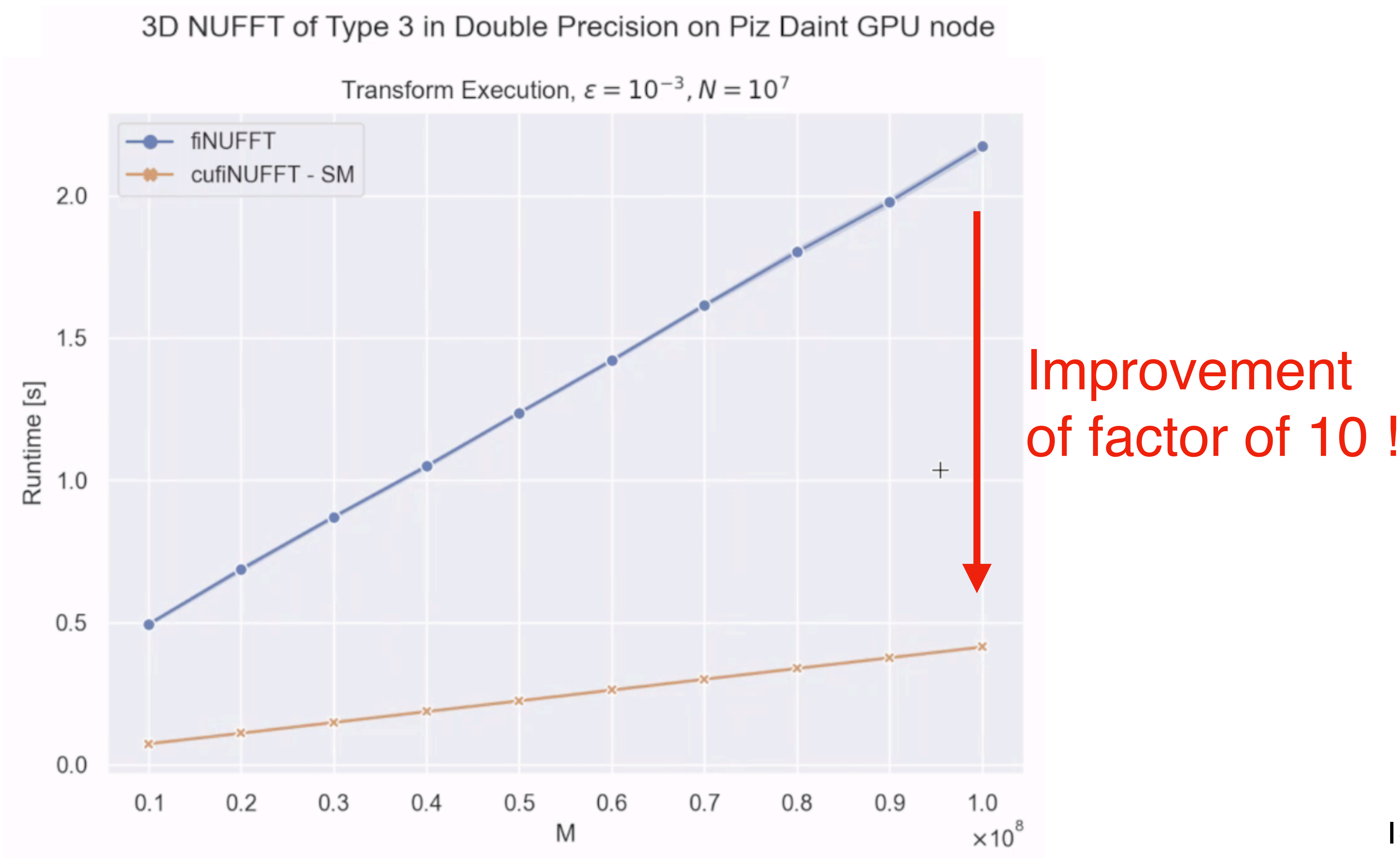


CLEAN (left) vs. Bluebild image (right) for the Toothbrush field. Contours of CLEAN are in red, contours of Bluebild in magenta, NVSS catalog overlaid (cyan squares).

Computational Imagine of interferometric data

From Emma Tolley (EPFL)

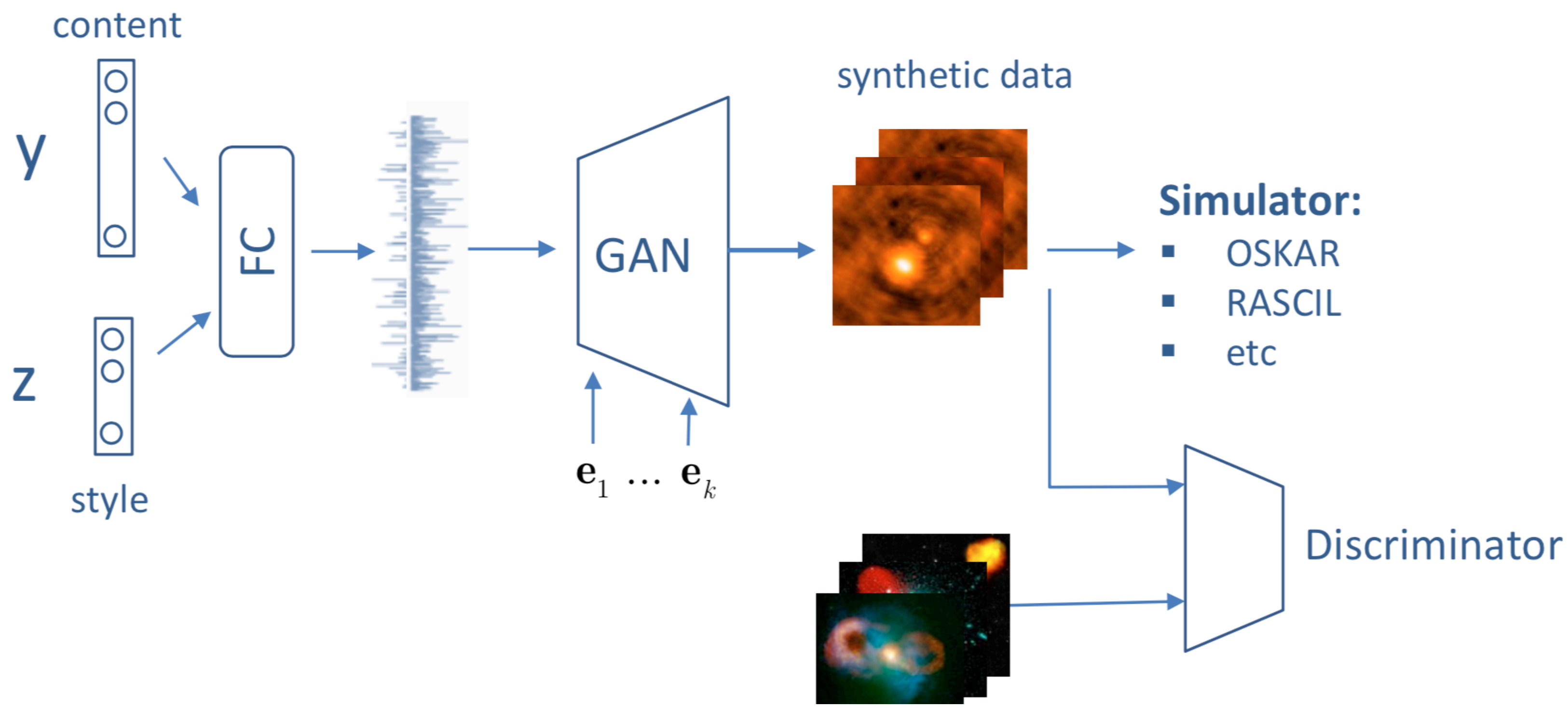
Developing: fast computational imaging codes that are more scalable than standard codes.



Innovative data analysis tools

From Slava Voloshynovskiy (UniGE)

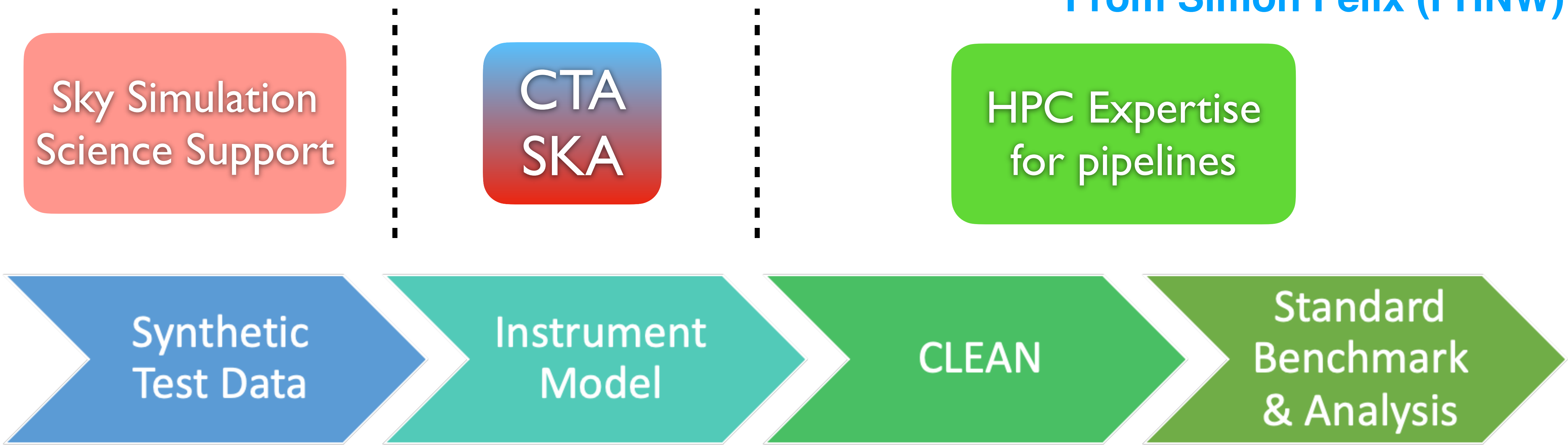
Example: Deep Conditional Generator to simulate any dataset (example for SDSS-like galaxy images)



Real data vs Synthetic data in red

Benchmarking new codes

From Simon Felix (FHNW)

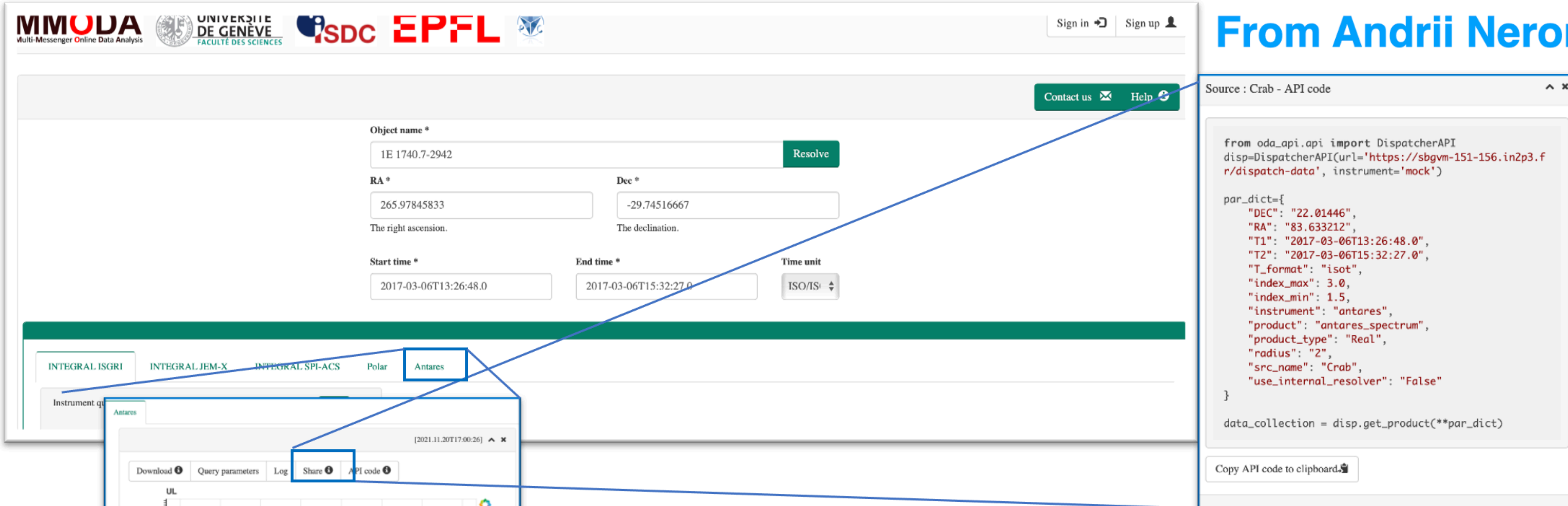


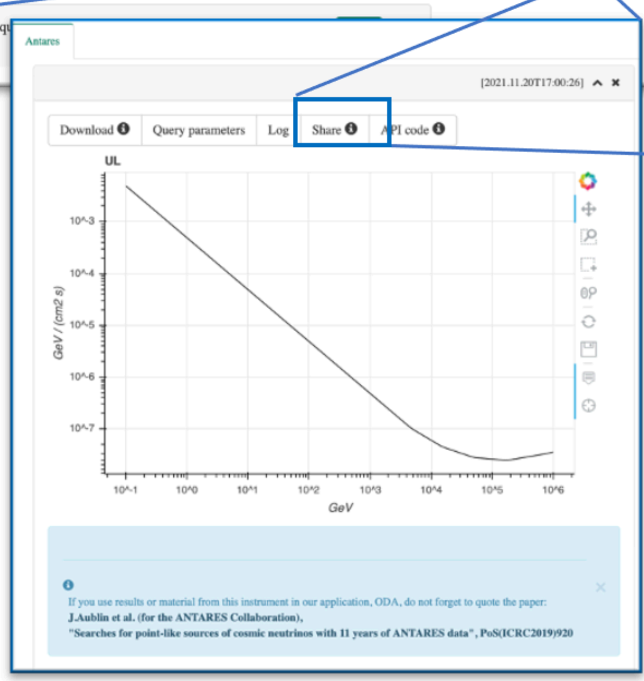
We provide pre-defined workflows

Anyone can replace individual parts

Online Data Analysis tools (ODA)

From Andrii Neronov (EPFL)





Online Analysis Platforms "as a service" (Workflow, notebooks, in web browser) will be used by *SKA Regional Center* and by the *CTA Science User Support System.*

Use FAIR principles *(Findable-Acessible-Interpoperable-Reusable) data management standard*

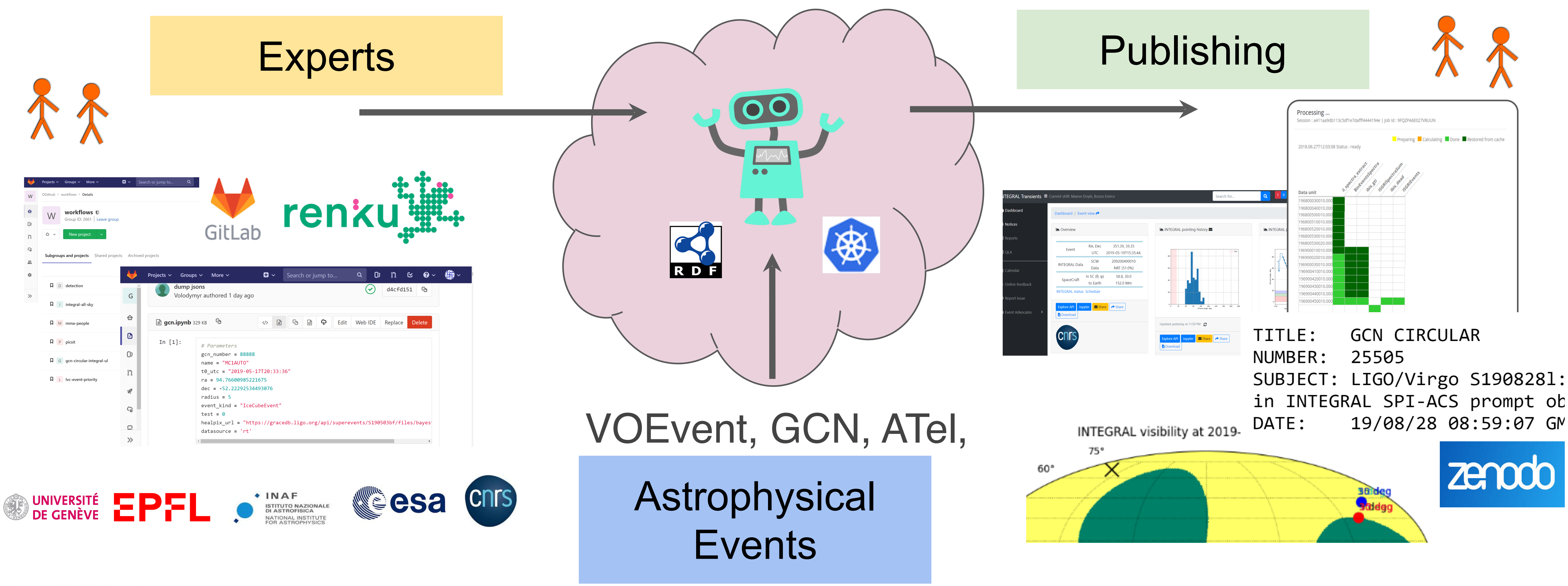
Examples:
1) **MMODA (Multi-Messenger Online Data Analysis platform)** developed by UniGE-EPFL-SDSC-APC-team,
2) ESCAPE Science Analysis Platform (ESAP) by H2020 Escape on CTA/SKA/KM3NET/VIRGO multi messenger,
3) DataLabs of ESA …

FAIR Astrophysical Transient Analysis in ODA

From Volodymyr Savchenko (UNIGE/EPFL)

Multi-messenger Transient Astronomy puts especially high demands on **confusion-free low-latency interoperability**, which we addressed by developing 2 key components:

- **Environment** to develop, test, and integrate data reduction, theoretical models, statistical methods, observatories operation tools.
- **Knowledge-Graph-enabled Engine** finds combinations of data, adapters, statistical methods, publishers, planning, and **disseminates provenance-tracked standard results** along with public data and code.



Conclusion

Many SKAO/CTAO synergy possibilities on various aspects!