

# FAIR data-intensive Science in SKA Regional Centres

Julián Garrido\*; Sebastián Luna; Susana Sánchez-Exposito; Lourdes Verdes-Montenegro; Michael Jones

Instituto de Astrofísica de Andalucía (CSIC)

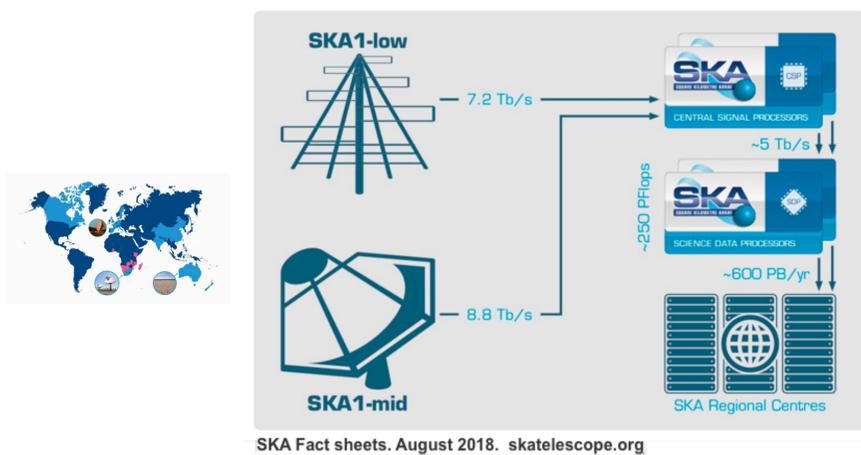
\* [jgarrido@iaa.es](mailto:jgarrido@iaa.es) ; <http://amiga.iaa.es>

We have long been aware of the need to face the challenge of handling SKA data to extract scientific knowledge, and our compromise is that this is done not only in an efficient way, but following the Scientific Method (hence in a reproducible way). Thus we are particularly engaged in the challenge of ensuring that Big Data science becomes Open Science in the SKA Regional Centres (SRCs), with the SRCs becoming a landmark not only in science and technology but also in Scientific Methodology. IAA-CSIC has started prototyping an SRC that will be fully aligned with the Open Science Principles.

Astronomy has a long-standing tradition of Open Access through the International Virtual Observatory Alliance (IVOA) but there are new solutions emerging from other fields, involving using e-Science technologies to enhance scientific collaboration, ensuring transparency, opening data collection and methods, or otherwise encouraging Open Science.

This work explains the workflow and steps that we have followed for conducting a scientific analysis in astronomy, following the Open Science and FAIR principles from its inception. It shows the Open Science European infrastructure providers selected and its connection with the IVOA ecosystem to support the whole cycle of FAIR research; and the scalability of this approach for an SRC.

## SKA



SKA is an international project to build the largest and most sensitive **radio telescope** ever conceived, being the greatest public data research project, once complete. It will be composed of thousands of antennas on Africa and Australia and it will generate a copious amount of data.

## Reproducibility in the Big Data era

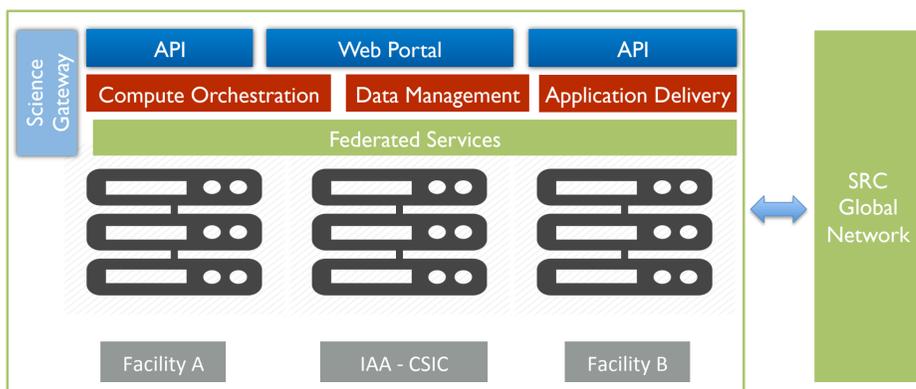
By definition, scientists want to follow the Scientific Method and therefore to ensure the reproducibility of their scientific results. However, something is failing when 70% of around 1500 researchers surveyed in a questionnaire on reproducibility answered that they have tried and failed to reproduce another scientist's experiments and >50% have failed to reproduce their own experiments (Baker, Nature, 2016). This means that **many scientists currently need additional support to achieve reproducibility with the current data volumes. Additional barriers will arise with such large data volumes as those generated by the SKA.** How will SRCs support astronomers to build reproducible methods?

## Fostering SRCs as Open Science Hubs

A worldwide distributed network of SKA Regional Centres (SRCs) will provide access to the SKA data, to analysis tools and processing power. The SRCs will be at the core of the exploitation of SKA data, being the place where the science will be done.

An Open Science implementation will facilitate sharing data, resources and tools across the global SKA community through the SRCs. Users will be able to verify, reuse and repurpose methods, accelerating discovery and the transfer of knowledge.

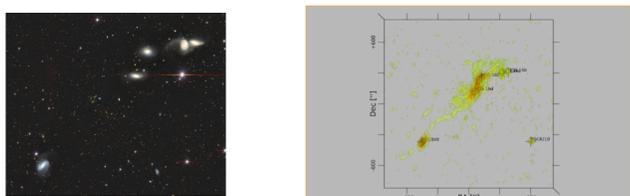
## Spanish SKA Regional Centre prototype



### Science Analysis Platform

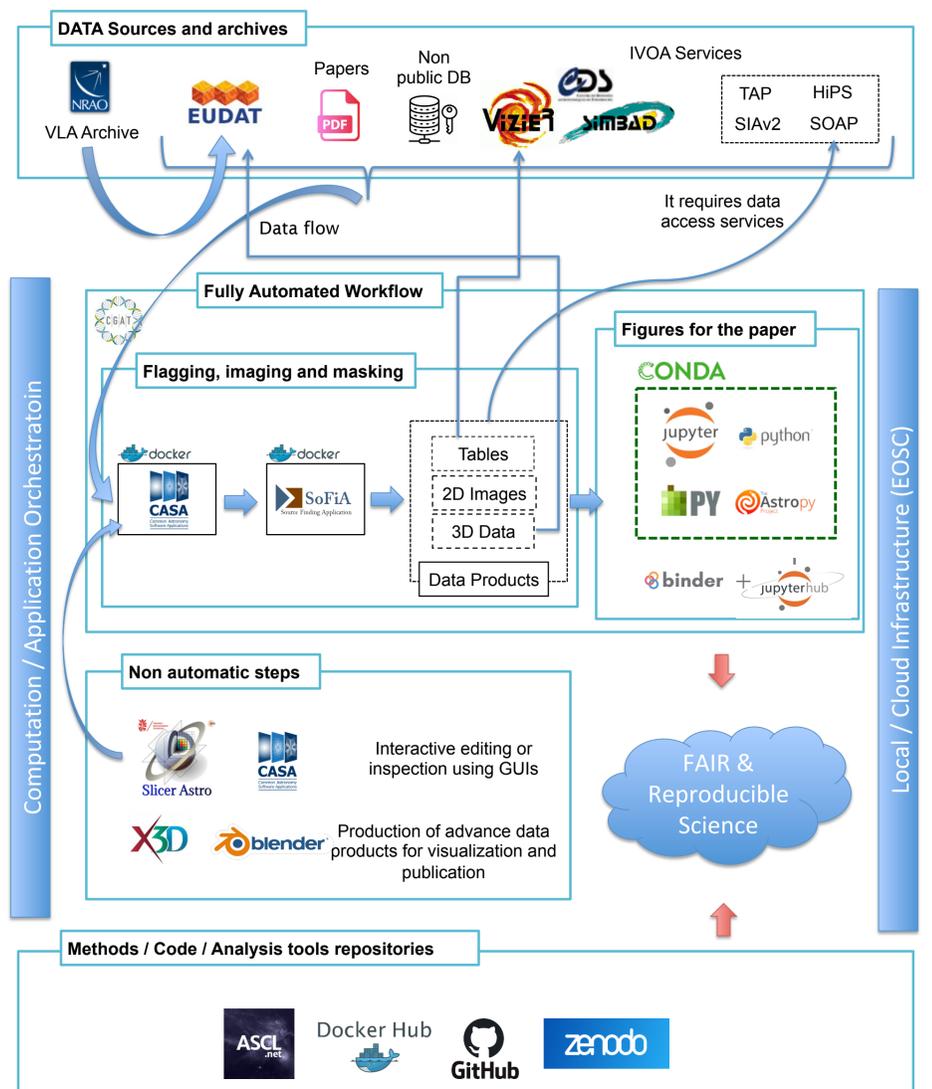
- Identify technical specifications from the use case requirements
- Set-up of the associated IAA computing/storage resources
- Partnership with national computing facilities
- Collaborations with e-Infrastructures and other SRCs prototype initiatives
- Provide a Science Gateway

## Case study



HCG 16 is complex compact group with starburst galaxies, AGN, tidal tails, etc. We have studied the HI content of the group and the on-going processes are causing it to change.

## Towards FAIR Science with existing tools



This work has been partly funded by projects AENEAS H2020 (731016), ESCAPE H2020 (824064), AYA2015-65973 (MINECO/FEDER, UE), RTI2018-096228-B-C31(MCIU/AEI/FEDER,UE).

