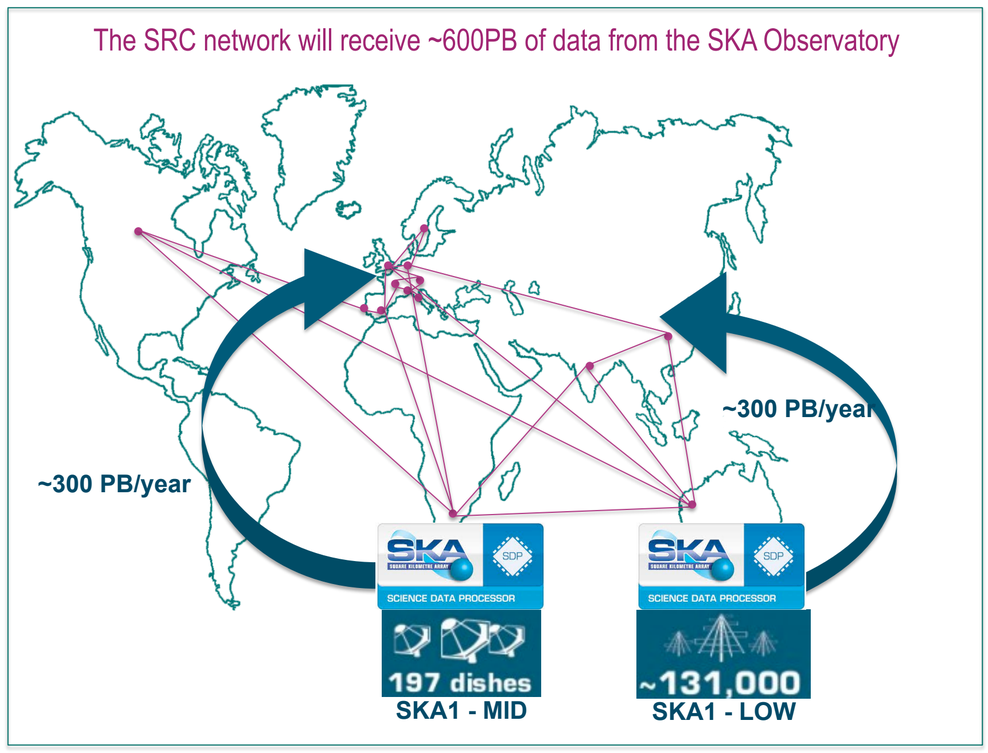
# SKA Regional Centre Prototype at IAA-CSIC: building an Open Science platform based on cloud services

Susana Sánchez-Expósito (sse@iaa.es); Sebastián Luna; Julián Garrido; Javier Moldón; Lourdes Verdes-Montenegro; Laura Darriba
Instituto de Astrofísica de Andalucía - CSIC. https://www.iaa.csic.es/

## Abstract:

The Square Kilometre Array (SKA) is a project to build a radio-interferometre, considered the largest generator of public data. Each year, it will produce around 600 Petabytes of data, which will be delivered to a network of SKA Regional Centres (SRCs). They will provide access to the SKA data and the analysis tools as well as the processing power necessary to fully exploit their science potential.

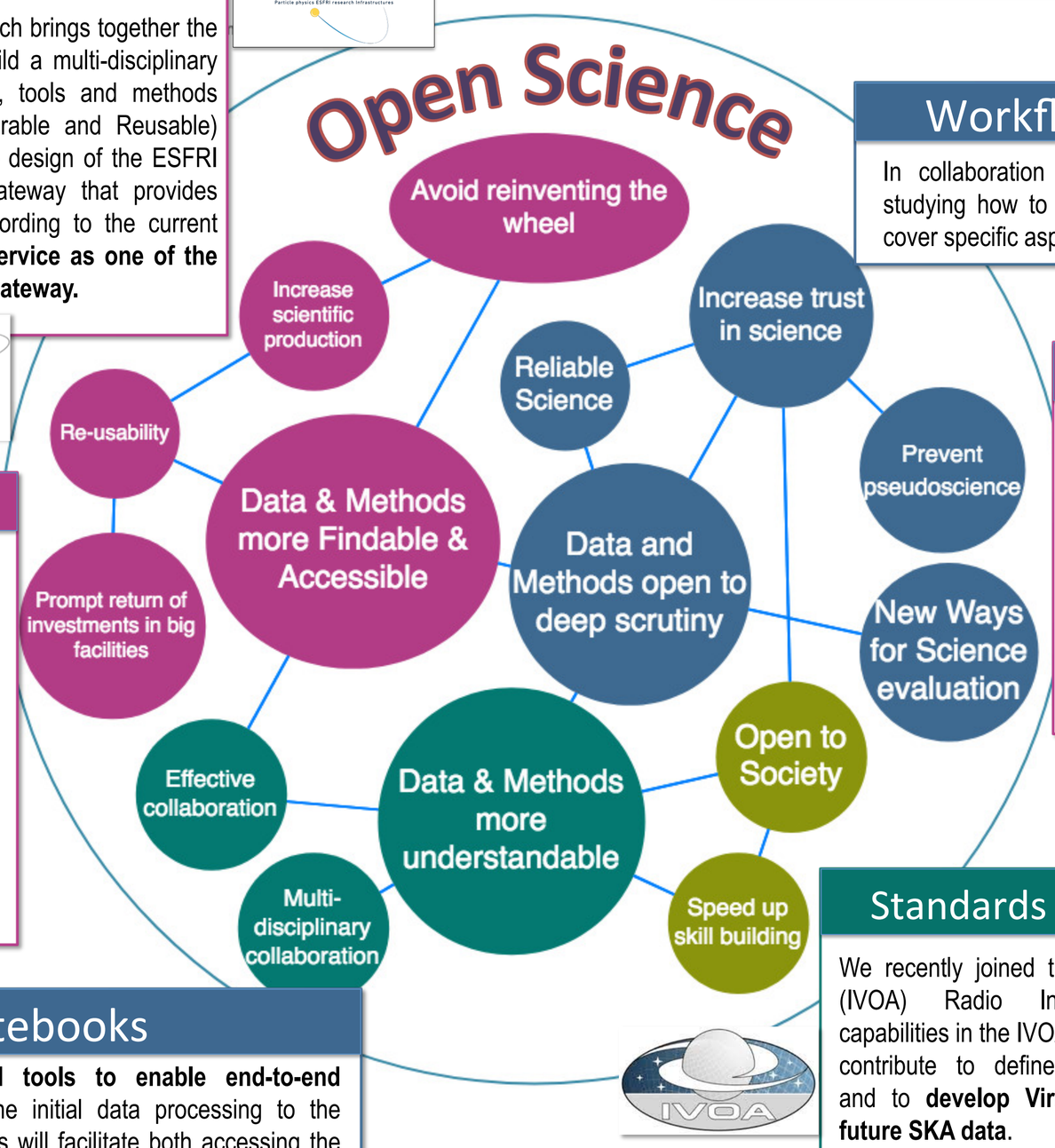
At the IAA-CSIC, in Granada (Spain) we are developing an SRC prototype aiming both to support users working with SKA precursors and pathfinders and to constitute a transversal, wavelength agnostic facility enabling knowledge exchange among a diverse community of users. We have deployed the first stage of the hardware, based on a cloud environment, and we are identifying and integrating components, such as a JupyterHub server or Virtual Observatory services, capable of supporting the scientific analysis according to the Open Science best practices.




Credits SKA logos: skatelescope.org

## Science Gateway

We are members of the H2020 ESCAPE project, which brings together the Astronomy and Particle Physics communities to build a multi-disciplinary open environment in which they can share data, tools and methods according to FAIR (Findable, Accessible, Interoperable and Reusable) principles. In this project, we are contributing to the design of the ESFRI Science Analysis Platform (ESAP), a Science Gateway that provides access to this open environment. In addition, according to the current ESAP design, **we have deployed a JupyterHub service as one of the components of the IAA-CSIC protoSRC Science Gateway.**




## Workflows and Provenance

In collaboration with the Spanish company GMV, we are studying how to adapt the IVOA **Provenance Data Model** to cover specific aspects of the SKA Science Data processor.

## Connected to EOSC

The user Authentication and Authorisation through the INDIGO Identity and Access Management service will **federate the IAA-CSIC protoSRC resources within the European Open Science Cloud**. We are carrying out this activity in the framework of the H2020 ESCAPE project and it will facilitate the integration of the prototype into the ESCAPE data lake.

## Science Archive

Key piece to enable scientific reproducibility, the design of the archive deserves special attention. Hence, we will contribute to the **Working Group, established by the SRC Steering Committee, in charge of the implementation and prototype of the SRC Science Archive,** chaired by the IAA-CSIC protoSRC coordinator (L. Verdes-Montenegro, co-chair: S. Gaudet -NRC).

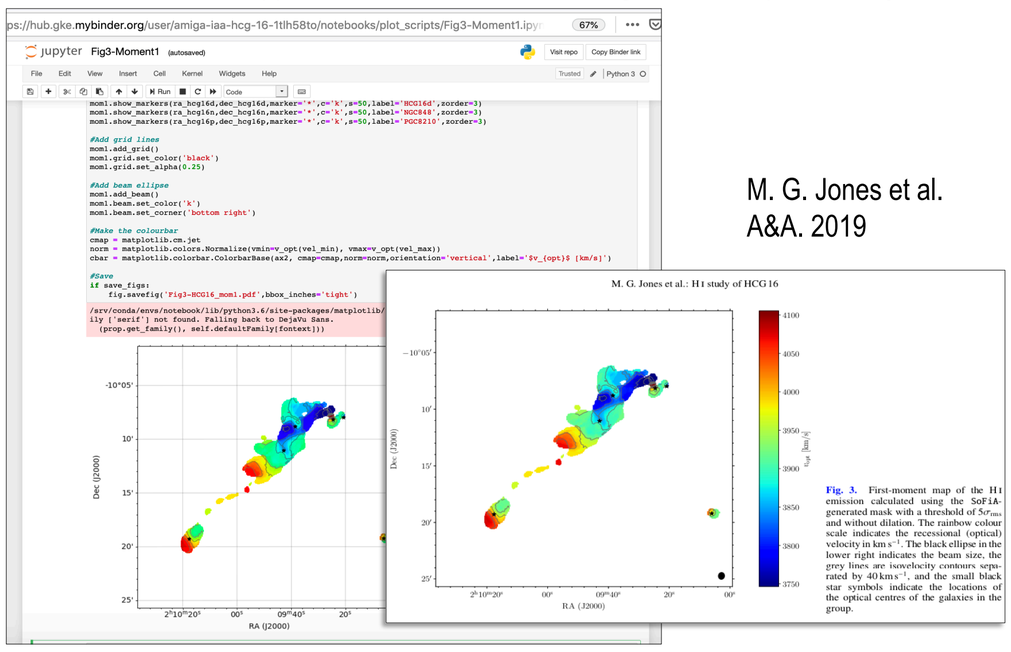
We are also participating in the development of a data archive for WALLABY, a survey to observe three-quarters of the whole sky in the 21-cm line of neutral hydrogen with the Australian SKA Pathfinder (ASKAP).

## Standards for data interoperability

We recently joined the International Virtual Observatory Alliance (IVOA) Radio Interested Group, aiming to develop radio capabilities in the IVOA ecosystem. In particular, our objective is to contribute to define standards for radio data interoperability and to **develop Virtual Observatory services for accessing future SKA data**.

## Reproducible Notebooks

**We are identifying a set of best practices and tools to enable end-to-end reproducibility of the scientific studies,** from the initial data processing to the generation of the plots and figures of the paper. This will facilitate both accessing the research data and exposing the underlying methods for a particular study as well as contributing to the verification of the analysis and the reusability of the digital experiment. In addition, we are leveraging the work that we did in the SKA-Link project ( http://amiga.iaa.es/p/330-SKA-Link.htm), whose main objective was to define a set of Best Practices, to be applied on the SRCs, for achieving scientific reproducibility.




M. G. Jones et al.
A&A. 2019

*Example of a figure of M.G Jones et al (2019) paper in which we have implemented best practices to achieve end-to-end reproducibility, here illustrated by showing how all figures of the paper can be re-generated in one-click by means of Jupyter Notebooks.*

## CLOUD PLATFORM - FLEXIBILITY

| | Front | Interior | Back |
|---|---|---|---|
| 42 | | openstack-sah | |
| 41 | | openstack-controller-01 | |
| 40 | | openstack-controller-02 | |
| 39 | | openstack-controller-03 | |
| 38 | | | |
| 37 | | openstack-nova-01 | |
| 36 | | openstack-nova-02 | |
| 35 | | openstack-nova-03 | |
| 34 | | openstack-nova-04 | |
| 33 | | openstack-nova-05 | |
| 32 | | | |
| 31 | | | |
| 30 | | | |
| 29 | | | |
| 28 | | | |
| 27 | | | |
| 26 | | 1GbE-S3048 | |
| 25 | | 100GbE-z9100-01 | |
| 24 | | 100GbE-z9100-02 | |
| 23 | | | |
| 22 | | ceph-mon-01 | |
| 21 | | ceph-mon-02 | |
| 20 | | ceph-mon-03 | |
| 19 | | | |
| 18 | | ceph-osd-08 | |
| 17 | | | |
| 16 | | ceph-osd-07 | |
| 15 | | | |
| 14 | | ceph-osd-06 | |
| 13 | | | |
| 12 | | ceph-osd-05 | |
| 11 | | | |
| 10 | | ceph-osd-04 | |
| 9 | | | |
| 8 | | ceph-osd-03 | |
| 7 | | | |
| 6 | | ceph-osd-02 | |
| 5 | | | |
| 4 | | ceph-osd-01 | |
| 3 | | | |
| 2 | | | |
| 1 | | | |

- OpenStack controllers
- Computing nodes
  - 2x 8TB SAS SSD
  - 384GB RAM (1 with 1TB)
  - 2x Intel Xeon 6230
- Admin network (1Gb)
- Computing network (100Gps)
- Ceph Monitors (Storage)
- Ceph OSDs (Storage)
  4:
  - 32x 8TB SAS SSD 12Gbps
  - 256GB RAM

  4:
  - 32x 4TB SAS SSD 12Gbps
  - 256GB RAM
- Connected to the Spanish Academic Network through a 10Gps link

The IAA-CSIC protoSRC will provide access to different cloud services:

SHORT TERM

**IaaS:** Virtual Machines (VMs) with different flavours to fulfil different RAM/disk/CPU needs.

**PaaS:** Customized VMs with specific software packages required for a scientific analysis.

LONG TERM

**SaaS**: From Virtual Observatory services for accessing the Science Archive to science domain software (e.g. CARTA viewer)

All authors of this work acknowledge financial support from the State Agency for Research of the Spanish MCIU through the "Center of Excellence Severo Ochoa" award to the Instituto de Astrofísica de Andalucía (SEV-2017-0709), as well as projects ESCAPE H2020 (824064), AYA2015-65973-C3-1-R (MINECO/ FEDER, UE), RTI2018-096228-B-C31(MCIU/AEI/FEDER,UE).