

Spanish Contribution to the Design and Development of the SRC Network v0.1

Susana Sánchez-Expósito, Julián Garrido, Lourdes Verdes-Montenegro, Laura Darriba, Manuel Parra-Royón, Javier Moldón, Jesús Sánchez, and María Ángeles Mendoza

Abstract – The SKA Regional Centre Network (SRCNet) will serve as the platform for scientific analysis and archiving of the data generated by the SKA Observatory (SKAO). Multiple international nodes work together to develop this platform, which will provide the astronomy community with access to SKAO data, as well as tools, computing resources, and scientific support. In this letter, we describe the main activities carried out by the Spanish team and how they contribute to the development of the SRCNet.

1. Introduction

The astronomical community is preparing for the operation of the SKA Observatory (SKAO). This observatory will consist of two radio interferometers located in Africa and Australia. These instruments will observe the universe at complementary frequencies, covering a broad range of the radio spectrum, and will provide a significant leap in sensitivity, resolution, and survey speed.

The SKAO will deliver calibrated data (~ 700 petabytes per year) to the SKA regional centers (SRCs), which are responsible for providing the scientific community with access to SKAO data and the processing and storage resources required for their scientific exploitation. The SRCs are independently resourced and collaborate with SKAO to build a global network of SRCs (SRCNet). The development of the SRCNet is transitioning from the prototyping phase to the implementation phase, getting ready to engage with telescope commissioning and science-verification activities.

Since 2011, the Instituto de Astrofísica de Andalucía (IAA-CSIC) has led Spain's scientific and technological

contributions to the SKA project. Highlights include publishing the Spanish SKA Science Book in 2015 and organizing key events such as “Synergies between J-PAS/J-PLUS and SKA Pathfinders in Preparation for SKA” and “The IAA-CSIC Severo Ochoa SKA Open Science School.” These efforts and the community's strong engagement resulted in Spain becoming an SKAO member in 2023. In 2018, the IAA-CSIC started the development of a Spanish SRC prototype [1] (the espSRC) to support preparatory activities for SKAO key science projects and facilitate scientific work with SKAO precursors and pathfinder telescopes while promoting best practices in open science.

In the following section, we provide more details about the computing platform and services provided by the espSRC. In Section 3, we explain the Spanish contribution to the SRCNet in the design (3.1) and prototyping (3.2) phases, as well as the position of the espSRC node in the first operational version of this network, the SRCNet v0.1 (3.3). Finally, a section with the conclusions is also included (4).

2. Computing Platform and Services Provided by the espSRC

The espSRC is designed to support a wide range of projects with diverse requirements, encompassing scientific studies that analyze data from different instruments, technological developments, and training in interferometric data processing and open science tools. We have implemented the espSRC as a cloud platform, leveraging the flexibility and elasticity of the cloud computing model to effectively accommodate this broad spectrum of projects. Since 2020, it has supported approximately 100 projects across research (i.e., projects for scientific data analysis), development (i.e., open science tools and support for the SRCNet), and training (i.e., support for schools). Among these projects, approximately 30% were dedicated to the development of the SRCNet (see section 3). Access to espSRC resources is generally granted upon submission of a science proposal outlining the project and technical needs. Acceptance criteria include alignment with SKA preparatory work and open science principles. Of note, the espSRC supported the Spanish team in the SKAO Science Data Challenge 2, where they achieved fifth place and received the Reproducibility Gold Award.

The espSRC platform currently comprises a cluster with six computing nodes, providing a total of 240 cores and eight storage nodes with a total raw storage capacity of 1.3 PB (data redundancies reduce this to an effective capacity of approximately 600 TB). On this cluster, we

Manuscript received 14 December 2024. This work was supported by grant CEX2021-001131-S funded by MICIU/AEI/10.13039/501100011033, grant PID2021-123930OB-C21 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, grant TED2021-130231B-I00 funded by MICIU/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR, grant RED2022-134464-T funded by MICIU/AEI/10.13039/501100011033, grant INFRA24023 funded by CSIC, the award “Center of Excellence Severo Ochoa” to the IAA-CSIC (SEV-2017-0709) funded by the AEI, grant EQC2019-005707-P funded by MICIU and the European Regional Development Funds, grant SOMM17-5208-IAA-201 funded by the Regional Government of Andalusia, and the coordination of the participation in SKA-SPAIN, funded by the Ministry of Science, Innovation and Universities (MICIU).

Susana Sánchez-Expósito, Julian Garrido, Lourdes Verdes-Montenegro, Laura Darriba, Manuel Parra-Royón, Javier Moldón, Jesús Sánchez, and Maria Angeles Mendoza are with Instituto de Astrofísica de Andalucía (Consejo Superior de Investigaciones Científicas), Glorieta de la Astronomía s/n, Granada, 18008, Spain; e-mail: sse@iaa.es, jgarrido@iaa.es, lourdes@iaa.es, ldarriba@iaa.es, mparra@iaa.es, jmoldon@iaa.es, jsanchez@iaa.es, amendoza@iaa.es.

have deployed OpenStack [2] to create and manage the espSRC cloud platform and Ceph [3] to implement a scalable storage system capable of providing a parallel file system, as well as block and object storage services. The espSRC cluster also includes three controller nodes, which are required to run various OpenStack services, and three monitoring nodes, which are essential for the Ceph system. All computing and storage nodes are interconnected via a 100 Gbps ethernet network.

The computing services offered by the espSRC are as follows:

- Virtual machines on-demand. A customizable cloud computing service using OpenStack, offering virtual machines and storage tailored to users' needs. Configurations (flavors) vary by CPU cores, RAM, network interconnection, storage, and operating systems.
- Data storage services. Two types of storage are offered, including (1) block storage, designed for intensive data processing and high-speed transfers, and (2) shared file system, optimized for data sharing across multiple virtual machines.
- Interactive processing with notebooks based on JupyterHub [4].
- Batch processing service based on SLURM [5].
- Container and software repository with ready-to-use astronomy software in containers to facilitate their deployment on the espSRC platform.

3. Spanish Contribution to the SRCNet

3.1 Design Phase

The SKA Regional Centre Steering Committee (SRCSC) established seven working groups (WGs) from 2020 to 2021 to gather requirements for the SRCNet and design its architecture. Each WG, co-chaired by two members of the SRCSC, was focused on the following specific domain: WG0 on the SRCNet architecture, WG1 on data logistics and transfers, WG2 on SRCNet operations, WG3 on federated computing and data services, WG4 on the SKAO science archive, WG5 on hardware requirements, and WG6 on engagement with the scientific community.

All SRCSC WGs included Spanish members, and WG4 was co-chaired by the Spanish representative in the SRCSC. A total of 14 participants from six different Spanish centers (the IAA-CSIC, the Barcelona Supercomputing Centre, the University of Valencia, the Institute of Physics of Cantabria, the Institute of Space Science, and the National Astronomical Observatory) contributed to the formal description of approximately 250 SRCNet requirements, which have been used in subsequent phases of the development.

3.2 Prototyping Phase

The SRCNet prototyping phase is a development project based on the SAFe agile methodology [6]. It began in April 2022, when seven international teams were established. One of these teams is the Coral team, which, until June 2024, includes the SRC teams from three different countries, Switzerland, Sweden, and Spain, as well as team members from the United Kingdom SRC initiative. The Spanish team plays an active role in coordinating the Coral team's activities, ensuring a focus on sustainability and open science since it contributes to the roles of scrum master and product owner in the agile methodology.

The Coral team is focused on building testbeds to evaluate SRCNet technologies. In particular, it has contributed to prototyping efforts by evaluating tools like Rucio [7] and OpenCADC Storage Inventory [8]. The team added international storage nodes, conducted performance tests, and provided feedback that informed the decision on which tool to adopt for SRCNet. A fundamental activity is the development of the MiniSRCNet, a multilayer prototype of the SRCNet platform, deployed across nodes in Sweden, Switzerland, Canada, and Spain. This demonstrator tested interoperability between core services, including data distributed system, science platform, authentication and authorization service, using real data from SKAO precursor telescopes. The Spanish team also designed tests to assess Rucio's data transfer scalability and performed benchmarks to optimize computing resource usage. These tests provided valuable insights into configurations for reliable, high-speed data transfers. All these activities aimed to emphasize sustainability, efficiency, and open science principles in the development of the SRCNet.

The Spanish team also contributed to gathering scientific workflows as use cases for SRCNet development. A software repository was created to share these workflows with SRCNet collaborators, along with a set of guidelines for adding new workflows. These guidelines ensure that all necessary elements for their re-execution (e.g., documentation, input data, configuration parameters, software dependencies, etc.) are included in the repository. Additionally, the team explored services to address federation at different levels, including creating a common interface for deploying applications across Kubernetes clusters using KubeFed and running cloud-native applications across multiple clusters.

Recently, the Architecture Core Team was established, including two members from the espSRC team, to support the SRCNet Architect in designing, implementing, and technically evaluating the software architecture of SRCNet. Its responsibilities include developing the SRCNet architecture roadmap, creating detailed architectural designs, and identifying areas for improvement.

3.3 Getting Ready for the SRCNet v0.1

In [9], the top-level roadmap of the SRCNet implementation is presented, covering different versions of the SRCNet until the first SRCNet public version, expected in

the first quarter of 2028, when the SKAO will be fully operational. This roadmap defines five versions of the SRCNet, aligning them with requirements arising from external activities outside the SRCNet context, such as SKAO construction milestones (e.g., completion of an array assembly) or scientific events. Version 0.1 will be the first operational version of the SRCNet, and it will be released in the first quarter of 2025, providing the following functionalities:

- Test data (synthetic data and data from SKAO precursor telescopes) can be disseminated among the different SRCNet nodes.
- Data can be accessed through a data lake (i.e., a distributed data system that works as a centralized repository capable of storing large amounts of structured and unstructured data).
- Data can be discovered through queries based on languages, such as structure query language or astronomical data query language.
- Data can be moved to an SRC node where local tools will facilitate basic data analysis and visualization.
- A unified authentication system is established for all the SRCNet nodes.

In February 2024, a call for expressions of interest was issued to participate in the SRCNet v0.1 production network and gather information about available resources to create a realistic implementation plan. The form requested responses regarding various requirements, including the software stack, human resources, network capabilities, storage, computing resources, and node validation tests. The espSRC was one of the nine nodes that replied. In March 2024, the first SRCNet workshop was held in Shanghai, where it was agreed to work on a scenario in which SRCNet v0.1 includes all SRCs that expressed an interest in contributing resources.

Starting in April 2024, the espSRC entered a procurement process for a storage upgrade. Its current infrastructure already meets the requirements to provide computing and networking resources for the SRCNet v0.1. The espSRC is fully engaged with the SRCNet

v0.1 implementation plan and has deployed all the compulsory services for the SRCNet v0.1.

4. Conclusions

The Spanish team has made a significant contribution to the development of the SRCNet during the design, prototyping, and development phases. Through the efforts of the IAA-CSIC and its collaborators, the espSRC has been established as a robust prototype, offering scalable computing and storage solutions tailored to the diverse needs of the SRCNet. A key example is the development of the MiniSRCNet prototype, which served to validate key technologies and their interoperability across the network, contributing to achieving seamless data management for scientific exploitation in the SRCNet. All the work during these phases has resulted in the espSRC becoming one of the nine nodes interoperating to build the SRCNet v0.1.

5. References

1. J. Garrido, L. Darriba, S. Sánchez-Expósito, M. Parra-Royón, J. Moldón, et al., “Toward a Spanish SKA Regional Centre Fully Engaged With Open Science,” *Journal of Astronomical Telescopes, Instruments, and Systems*, **8**, 1, January 2022, p. 011004.
2. OpenStack Foundation, “OpenStack: Open Source Cloud Computing Infrastructure,” <https://www.openstack.org> (Accessed 7 January 2025).
3. Ceph Foundation, “Ceph: Open Source Distributed Storage,” <https://ceph.io> (Accessed 7 January 2025).
4. Jupyter, “Project Jupyter,” <https://jupyter.org> (Accessed 7 January 2025).
5. SchedMD, “Slurm: Simple Linux Utility for Resource Management,” <https://slurm.schedmd.com> (Accessed 7 January 2025).
6. Scaled Agile, Inc, “SAFe Framework,” <https://scaledagileframework.com> (Accessed 7 January 2025).
7. M. Barisits, T. Beermann, F. Berghaus, B. Bockelman, J. Bogado, et al., “Rucio: Scientific Data Management,” *Computing and Software for Big Science*, **3**, 2019, pp. 11.
8. Canadian Astronomy Data Centre, “Storage Inventory Architecture and Deployment,” <https://www.opencadc.org> (Accessed 12 December 2024).
9. SKA Observatory, “The SRCNet Top-Level Roadmap,” <https://confluence.skatelescope.org/download/attachments/74711238/SRCNet%20Top%20Level%20Roadmap.pdf> (Accessed 12 December 2024).