Highlights of Spanish Astrophysics XII, Proceedings of the XVI Scientific Meeting of the Spanish Astronomical Society held on July 15 - 19, 2024, in Granada, Spain. M. Manteiga, F. González Galindo, A. Labiano Ortega, M. Martínez González, N. Rea, M. Romero Gómez, A. Ulla Miguel, G. Yepes, C. Rodríguez López, A. Gómez García and C. Dafonte (eds.), 2025

Spanish contribution to the development of the SKA Regional Centres Network: A sustainable approach

Sánchez-Expósito, S.[1], Verdes-Montenegro, L.[1], J. Garrido[1], L. Darriba[1], M. Parra[1], J. Moldón[1], J. Sánchez[1], M.A. Mendoza[1], and A. García[1].

[1] Instituto de Astrofísica de Andalucía (Consejo Superior de Investigaciones Científicas)

Abstract

The SKA Observatory (SKAO) has started to install its first antennas. In total, SKAO will consist of 130.000 low-frequency dipoles in Australia (SKA-Low) and 197 dishes in South Africa (SKA-Mid). By January 2025, the installation of the first set of antennas (AA0.5) will be finished and the process of science verification will start in 2026. Very soon, the unprecedented sensitivity of SKAO will help the community to address fundamental questions in Astrophysics, Fundamental Physics and Astrobiology.

SKAO data ($\sim$ 700 Petabytes/year) will be delivered to a network of distributed SKA Regional Centres (SRCs), which are in charge of providing the community with access to the SKAO data as well as to the tools and computing resources required for their scientific analysis. The SRC Network (SRCNet) is being designed to ensure a seamless access to the different computing resources available in the SRCs, applying the "moving code to the data" paradigm and allowing analytical tasks executed in any SRC, regardless of where the user submitted them.

The development of the SRCNet is moving from the prototyping phase to the implementation phase, getting ready to engage with the telescope commissioning and the science verification activities.

In this paper we will present the Spanish contribution to the development of the SRCNet. This contribution is characterised by our interest in applying Sustainability and Open Science principles to enhance the accessibility and reusability of the SKAO data and methods. Among other activities, we will describe the implementation of a SRCNet demonstrator that integrates the computing resources from 3 SRCs. This demonstrator is considered key for the evaluation of technologies to be used in the SRCNet implementation. We will also present how we are addressing the challenge of improving the energy efficiency at the SRCs.

1 Introduction

The SKA Observatory (SKAO) will have two radio interferometers located in Africa and Australia, which will observe the Universe in complementary frequencies, covering a wide range of the radio spectrum. The initiative to build this facility is an "European Strategy Forum on Research Infrastructures" (ESFRI) landmark project. SKAO will provide a significant leap in sensitivity, resolution and survey speed, which will enable groundbreaking scientific discoveries in fields such as Astrophysics, Astrobiology and Fundamental Physics.

The SKAO will be the largest generator of public data. The construction of the SKAO is expected to be completed in 2029, with scientific verification scheduled for 2026. Collaborations for preparing the SKAO Key Science Projects (KSPs) will form between 2027 and 2029. The SKAO will deliver calibrated data (~ 700 petabytes per year) to an international network of SKA Regional Centres (SRCs), known as the global SRC Network (SRCNet).

Currently, sixteen countries have an SRC initiative. They are collaborating with SKAO to develop the SRCNet as a seamless infrastructure to provide the scientific community with access to SKAO data, as well as to the computing resources and tools needed for its scientific exploitation. The SRCs will thus serve as the scientific core of the SKAO, with their prototypes acting as a natural space for national communities to prepare for this challenge by utilizing SKAO precursor telescopes.

The SRCNet project began its prototyping phase in April 2022 and is currently preparing to transition to the implementation phase, scheduled for the first quarter of 2025 [2]. During this phase, the SRCNet will participate in telescope commissioning and science verification activities.

Spain joined the SKA Observatory as a member in 2023. The Instituto de Astrofísica de Andalucía (IAA-CSIC) is the coordinator of the Spanish scientific and technological efforts related to the SKAO initiative. In 2018, the IAA-CSIC, with the support of its Severo Ochoa programme, started to develop the Spanish SRC prototype (espSRC) [1] to support preparatory work for SKAO Key Science Projects and scientific engagement with SKAO precursor and pathfinder telescopes, while also fostering best practices in Open Science. The team building the espSRC is contributing actively to the SRCNet development, placing a special emphasis on applying Sustainability and Open Science principles.

The next sections further explain the espSRC initiative (section 2) and its contribution to the SRCNet development (section 3) including the coordination of development activities (section 3.1), the contribution to the Data Management system prototypes (section 3.2), to the MiniSRCNet demonstrator (section 3.3), and to the execution of tests and benchmarks to evaluate the performance of the services. At the end, a section with the conclusions is also included (section 4).

2 The Spanish SRC prototype initiative

The initiative to develop a Spanish SRC prototype started in 2018 with the support from the Severo Ochoa programme of the IAA-CSIC. The main goals of this initiative are:

- Enable the community to extract the utmost scientific value from the SKAO.
- Maximise participation in KSPs through dedicated support.
- Facilitate preparatory SKAO science with SKAO precursors and pathfinders.
- Support the community to acquire the scientific and technical skills for the new SKAO paradigm and for the Open Science.
- Reinforce Multifrequency / Multimessenger synergies.

To achieve these objectives, this initiative has set up a computing infrastructure with the tools and services required by the community to carry out its scientific studies. It also provides user support for radio data processing and the use of Open Science tools, as well as organises training events on those topics.

The espSRC computing infrastructure is a cloud platform that currently has 6 compute nodes, 8 storage nodes, 3 control nodes and 3 monitoring nodes. This setup provides a total of 240 physical cores, 2.8 Terabytes of memory, and a raw storage capacity of 1.3 Petabytes approximately. All cluster nodes are interconnected via a 100 Gbps Ethernet network.

Thanks to the flexibility and scalability of the cloud computing model, the espSRC is able to support a wide range of projects, offering different services to the community. The main ones are listed below:

- Virtual Machine service.
- Disk storage.
- Interactive analysis service based on JupyterHub.
- Batch processing service based on Slurm.
- Software and container catalog.

Since the beginning of its operations in early 2021 and until June 2024, the espSRC has supported 86 projects: 31 focused on scientific studies, 42 dedicated to the development of new services or the contribution to the SRCNet (see next section) and 13 for supporting training events. The research project supported are mainly focused on the processing of data from radio telescopes, but also scientific studies from other astrophysics fields are supported as shown in Fig. 1.

3 Spanish Contribution to the SRCNet

3.1 Contribution to the coordination of development activities

The SRCNet follow the SAFe methdology that implement an iterative development with agile practices that allow continuous feedback and adaptation based on the user requirements. In

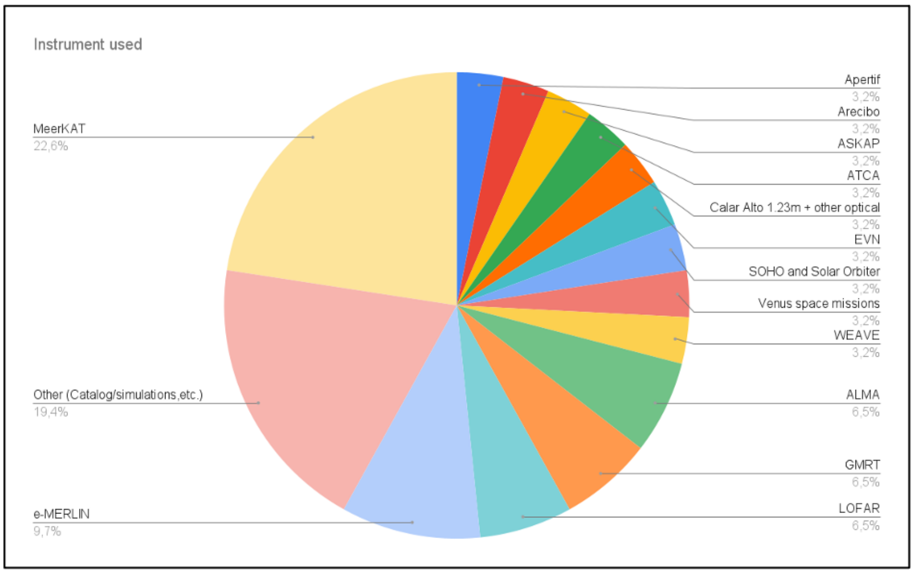

Figure 1: Pie chart showing research projects supported by the espSRC, categorised by the primary instrument they focus on.

the SRCNet project, the development is broken into smaller increments that will be finished in 3 months. SRCNet developers work together in agile teams that include the Scrum Master and Product Owner role. The Scrum Master is the team member that helps the team to improve on their delivery and is focused on removing obstacles for flow in the team. The Product Onwer is the responsible of defining tasks for the team and prioritizing them, making sure that the objectives that those tasks contribute to are fulfilled.

In April 2022 7 agile teams were defined for the SRCNet development and members of the espSRC team were included in the one called Coral. This team was formed by gathering the contribution from 3 different SRC initiatives: Sweden, Switzerland and Spain, together with a member from the United Kingdom. Coral is the largest SRCNet agile team and it is aimed at building testbeds for evaluating technologies considered to be integrated in the SRCNet.

The espSRC contributes with the Product Owner and Scrum Master roles, and this fact allowed focusing several Coral activities on developments related to the Open Science and the sustainability of the SRCs. Next sections describe these activities.

3.2 Contribution to the Distributed Data Management prototypes

The Distributed Data Management (DDM) system will be in charge of the data transfers, the access to the data and data replication. Two different tools were considered to implement it: Rucio, developed by CERN, which is able to manage large datasets stored in globally distributed and heterogeneous storage backends; OpenCADC Storage Inventory, a tool developed by the Canadian Astronomy Data Centre (CADC) to manage archival file storage.

The Coral team contributed significantly to the development of those prototypes by adding

3 international storage nodes to them, providing feedback, developing documentation and supporting other teams to integrate their resources into those prototypes. The Coral team also contributed to the decision process to select one of these two technologies, executing tests and providing information about their performance and the operational capacity.

3.3 Contribution to the MiniSRCNet demonstrator

The Mini-SRCNet is a scaled-down version of the SRCNet platform for testing. The Coral team, with the support from the Canadian SRC team, deployed this demonstrator in 4 nodes (Sweden, Switzerland, Canada, Spain). Real data from SKAO precursor telescopes was distributed by the team to evaluate the system performance. In this prototype several SRCNet modules inter-operated: the DDM system, the Authentication and Authorization service, and the science analysis platform, providing very valuable feedback on the internal interfaces of the SRCNet.

3.4 Tests and Benchmarks

The Spanish members of the Coral team designed several tests to study Rucio's data transfers, in particular, its scalability with large datasets and high transfer volumes. Those tests also examine how different RSE access protocols impact transfer efficiency and assesses the performance of the File Transfer Service (FTS) under varying loads. The aim was to identify any limitations or bottlenecks in Rucio's transfer capabilities and provide insight into optimal configurations for reliable, high-speed data transfers. Those tests were complemented with a network test performed between a Data Transfer Node in London and the espSRC platform.

Fig. 2 presents the results of one of these tests, showing how data transfer performance improves as the size of the data file increases.

The Coral team also evaluated different benchmarking tools (e.g. the Roof model), which are capable of providing information on how the scientific codes use the computing resources (memory and CPU). This information could be very valuable to identify how to improve the codes so that they make a better use of the resources and also in the other way round, how to select the hardware that fulfil the requirements of these codes.

4 Conclusions

The SRCNet is fundamental for achieving the goals of SKAO. It will be the platfom where the community will analyze SKAO data. Building the SRCNet requires a substantial effort, since it must integrate independently funded and technologically heterogeneous nodes to provide seamless access to SKAO datasets and analytical tools. The IAA-CSIC is leading the development of the espSRC, an initiative that supports the national community while contributing to the international SRCNet project. Since 2021, the espSRC has supported over 80 projects. The espSRC team has contributed to create several testbeds, tests and benchmarks, which have provided valuable insights to evaluate the technologies considered for the SRCNet. Throughout this process, the team has emphasised Open Science and efficiency, contributing to a more sustainable SRCNet.

average transfer rate as a function of RSE and file size

file size \ RSE	SPSRC_STORM	KRSRC_STORM	JPSRC_STORM	IMPERIAL	CNSRC_STORM	CNAF	CHSRC_XRD	CASRC_XRD
200 MB	2.30 MB/s	1.65 MB/s	1.76 MB/s	2.14 MB/s	1.65 MB/s	2.09 MB/s	2.17 MB/s	1.71 MB/s
500 MB	5.73 MB/s	3.52 MB/s	3.73 MB/s	5.11 MB/s	3.38 MB/s	4.97 MB/s	5.20 MB/s	3.72 MB/s
1 GB	10.6 MB/s	4.83 MB/s	5.25 MB/s	8.56 MB/s	4.60 MB/s	8.96 MB/s	8.84 MB/s	5.41 MB/s
1.50 GB	15.7 MB/s	6.75 MB/s	7.23 MB/s	12.3 MB/s	6.15 MB/s	13.1 MB/s	12.8 MB/s	7.50 MB/s
2 GB	20.6 MB/s	7.58 MB/s	8.09 MB/s	15.1 MB/s	6.93 MB/s	16.0 MB/s	15.8 MB/s	8.61 MB/s
3 GB	28.9 MB/s	8.71 MB/s	9.43 MB/s	19.7 MB/s	7.83 MB/s	20.8 MB/s	20.5 MB/s	9.92 MB/s
Mean	14.0 MB/s	5.51 MB/s	5.92 MB/s	10.5 MB/s	5.09 MB/s	11.0 MB/s	10.9 MB/s	6.15 MB/s

Figure 2: Results of executing a Rucio data transfer test, in which files of different sizes (i.e. 200, 500, 1000,1500, 2000, 3000 Megabytes) were transferred between the different nodes of the Rucio Data Management system (i.e. SPSRC_STORM, KRSRC_STORM, JPSRC_STORM, IMPERIAL, CNSRC_STORM, CNAF, CHSRC_XRD, CASRC_XRD).

Acknowledgments

The authors acknowledge financial support from the grant CEX2021-001131-S funded by MICIU/AEI/ 10.13039/501100011033, from the grant PID2021-123930OB-C21 funded by MICIU/AEI/10.13039/ 501100011033 and by ERDF/EU, from the grant RED2022-134464-T funded by MICIU/AEI/ 10.13039/ 501100011033, from the grant TED2021-130231B-I00 funded by MICIU/AEI/ 10.13039/ 501100011033 and by the European Union NextGenerationEU/PRTR, from The coordination of the participation in SKA-SPAIN, funded by the Ministry of Science, Innovation and Universities (MICIU), from the grant INFRA24023 funded by CSIC, and the coordination of the participation in SKA-SPAIN, funded by the Ministry of Science, Innovation and Universities (MICIU) and from the grant PTA2022-022026-I funded by MICIU/AEI/ 10.13039/501100011033 and by the ESF+.

The authors acknowledge the Spanish Prototype of an SRC (SPSRC) service and support funded by the Ministerio de Ciencia, Innovación y Universidades (MICIU), by the Junta de Andalucía, by the European Regional Development Funds (ERDF) and by the European Union NextGenerationEU/PRTR. The SPSRC acknowledges financial support from the Agencia Estatal de Investigación (AEI) through the "Centre of Excellence Severo Ochoa" award to the Instituto de Astrofísica de Andalucía (IAA-CSIC) (SEV-2017-0709) and from the grant CEX2021-001131-S funded by MICIU/AEI/10.13039/501100011033.

References

[1] Garrido, J., Darriba, L., Sánchez-Expósito, S., Parra-Royón, M., Moldón, J., Mendoza, M. Á., Luna-Valero, S., Alberdi, A., Márquez, I., & Verdes-Montenegro, L. 2022, Journal of Astronomical Telescopes, Instruments, and Systems, 8. doi:10.1117/1.JATIS.8.1.011004

[2] Salgado, J., Bolton, R., Swinbank, J., Joshi, R., Sánchez, S., Villote, J.-P., Gaudet, S., Yates, J., Barbosa, D., Taffoni, G., Frank, B., van Haarlem, M., Breen, S., Conway, J., Akahori, T., Yates, J., Tolley, E. E., Wadadekar, Y., Lee-Waddell, K., de Boer, J., 2023, SRC-0000002