A CANADIAN SQUARE KILOMETRE ARRAY REGIONAL CENTRE

Kristine Spekkens (RMC; Canadian SKA Science Director), Erik Rosolowsky (University of Alberta), Séverin Gaudet (NRC), Michael Rupen (NRC) and the Association of Canadian Universities for Research in Astronomy (ACURA) Advisory Council on the SKA (AACS)

This white paper submitted in response to NDRIO's first call for input on Canada's future (digital research infrastructure) DRI ecosystem presents the prospects for a Canadian Square Kilometre Array (SKA) Regional Centre (SRC). The SKA is a global radio astronomy observatory that will make transformational advances in our understanding of the universe across a wide range of fields, ranking among the highest infrastructure priorities for the Canadian astronomical community in the coming decade. The enormous data rates anticipated for the SKA imply that its scientific exploitation will require significant storage and computing resources provided by a global network of SRCs. A Canadian SRC – the only such facility in the Americas – would leverage our strength in computing platform and archive development to serve the scientific data processing, storage and user support needs of Canadian SKA users. We advocate for a partnership between NDRIO, CANARIE, the CADC, CANFAR and universities to operate a Canadian SKA Regional Centre as a hub of SKA digital innovation that integrates seamlessly and synergistically into the broader national DRI ecosystem.

1 Introduction: Canada and the Square Kilometre Array

The Square Kilometre Array (SKA) is a global radio observatory that will enable transformational science about the history, contents, extreme conditions, and prospects for life in the Universe. The first phase of the project, called SKA1[1], will be one of the largest scientific endeavours in history, involving the interconnection of hundreds of radio dishes and thousands of radio antennas spread across two continents, and drawing on partnerships between astronomers, computer scientists, and engineers from almost two dozen partner countries, including Canada. The SKA facilities are international (see Figure 1): the project is headquartered in the United Kingdom, while the dishes are located in South Africa, the antennas in Australia, and the scientific data centres around the world. Construction is set to begin in 2021, with full operations starting in 2028. The SKA, operating from 50 MHz to 15 GHz, will be the largest and most powerful general-purpose radio telescope for years to come, monitoring the radio sky in unprecedented detail.

Canada has a long history of scientific and technological engagement with the SKA, and has made important contributions on both fronts in the decade-long SKA design phase (Spekkens+ 19; http://skatelescope.ca). Canada is now poised to play leadership roles in SKA science and technology during construction and operations. The scientific goals of the SKA align well with the strengths of Canadian astronomers, namely our world leadership in the study of pulsars, cosmic magnetism, the bursting radio sky, and low-frequency cosmology, along with our multi-wavelength expertise in galaxy evolution, multi-messenger astronomy, and planetary system formation. Key Canadian SKA technological capabilities include the design and fabrication of digital signal processors, low-noise amplifiers and digitisers. Following an in-depth nationwide review, the 2020–2030 Long Range Plan for Astronomy and Astrophysics (Barmby, Gaensler+ 20) ranks the SKA as a top infrastructure priority for Canadian researchers in the next decade.

The raw data rates and calibrated data volumes produced by the SKA will be enormous. The facilities in Australia and South Africa will stream 7.2 TB/s and 8.8 TB/s of raw data into raw signal processors, respectively.

[1]This white paper considers exclusively the SKA1 facility and its DRI requirements as defined in Dewdney+ 16 and Quinn+ 20. For the purposes of this document, SKA = SKA1.

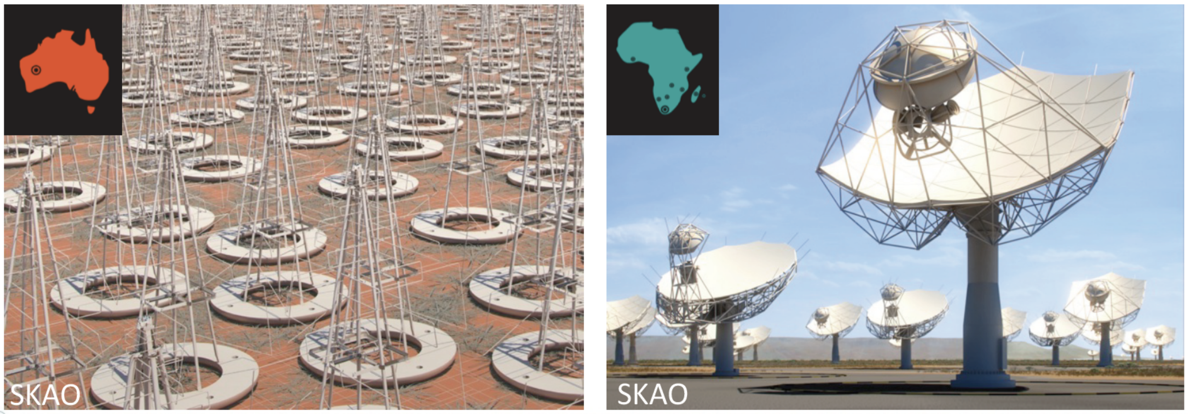

Figure 1: Artist's conception of the SKA, with the lower-frequency radio antennas in Australia (left) and higher-frequency radio dishes (right) in South Africa. Their approximate locations on those continents is inset, with long baseline interferometry stations across the African continent also shown on the left. Image credit: SKAO.

These in turn will feed ~ 5 TB/s into science data processors, that will use ~ 250 Pflops to produce ~ 600 PB/yr of calibrated data products (Quinn+ 20). A network of data centres called SKA Regional Centres (SRCs) will handle the global science processing, archive and user support needs, following a tiered model like that adopted by CERN (the European Organization for Nuclear Research). The SKA is the most compute- and storage-intensive facility prioritized in the 2020 Long Range Plan, and the corresponding SRC DRI requirements for Canada are listed separately therein. Since the SRC is a vital component of Canadian participation in the SKA, the costs of the SRC will be requested as part of the broader request to Government for SKA funding in the next decade.

This White Paper focuses on the DRI associated with a Canadian SRC in the NDRIO context. It is complementary to the white papers submitted by the Canadian Astronomical Society's Computation and Data Committee (CDC, Lovekin + 20) and the NRC-operated Canadian Astronomy Data Centre (CADC, Kavelaars+ 20). The 2020 Long Range Plan has also been submitted to this call as a reference. In the current paper, Section 2 describes the SRC Model, related DRI requirements and costing of a Canadian SRC, and Section 3 advocates for a partnership between NDRIO, CANARIE, the CADC, the Canadian Advanced Network for Astronomical Research (CANFAR) and Canadian universities to deliver these resources within the future DRI landscape.

2 A Canadian SKA Regional (Data) Centre

At the onset of full operations towards the end of this decade, the SKA will generate $\sim$ 600 PB of calibrated science data products each year. This data rate is unprecedented in observational astronomy. The infrastructure for transporting such large data volumes to users around the world, and the computational resources that are required to enable users to turn those data into an additional $\sim$ 100 PB per year of scientific results, will be provided by a collaborative interoperating network of SKA Regional Centres (SRCs). Users connect to these SRCs through a Science Gateway that is managed by the SRC Alliance (see Figure 2 left). Approximately 5 SRCs will be needed to handle the global scientific DRI requirement for the SKA.

A minimum set of requirements for each individual SRC will be defined by the SKA Observatory through a number of international working groups to which Canada contributes. These requirements will be related to:

- the curation and preservation of SKA- and user-generated data products and workflows;
- the provision of resources for post-processing, analysis, and data visualisation;
- the application of SKA data policies and procedures for access to SKA data; and

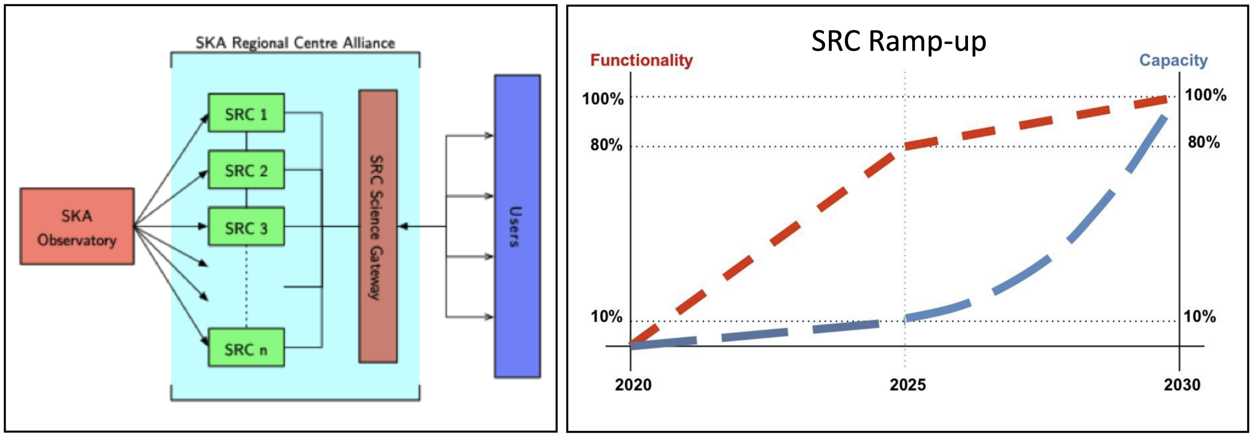

Figure 2: *Left:* A global network of SKA Regional Centres (SRCs) receives science data products from the SKA Observatory. Access to SKA science data products, as well as the tools and processing power necessary to fully exploit the science potential of these products, is provided by a Science Gateway with which users from around the world interact. The 2020 Canadian Long-Range Plan for Astronomy and Astrophysics recommends that Canada host an SRC in the network, one of 5 such facilities around the world. *Right:* Ramp-up of SRC functionality (red, corresponding to personnel) and capacity (blue, corresponding to storage/processing) required through 2030. The corresponding resources requirements and cost estimates are given in Figure 3. Image credit: Quinn+ 20.

- the provision of user support.

In addition, many SRCs (including that advocated for Canada) will have components for education and public outreach (EPO) and to enhance equity, diversity, and inclusivity (EDI).

The SRC Alliance that oversees the network of SRCs will have two primary, collective, responsibilities: (i) to provide the long-term curation and preservation of SKA and user-generated data products and workflows; and (ii) to provide a common platform to enable the creation of user-generated data products of high scientific value. The aim is to create an SRC Alliance working together for the SKA community as a whole. As with most modern, large scientific projects, the SKA will execute complex projects over a broad and globally distributed community with common aims. The following minimum functionality for the SRC Alliance is anticipated to be required:

- provision of a long-term data archive for SKA- and user-generated science data products (known as Observatory data products and advanced data products, respectively);
- sharing of data products, processing resources, and workflows between SRCs;
- interoperability, using a common platform and standards, between SRCs;
- interoperability with other astronomical facilities, through the support of Virtual Observatory services and protocols.

The DRI resources that Canada would need to contribute to the SRC network from 2021 – 2024 (the period specified in the NDRIO white paper call) and 2021 – 2030 (the period considered in the 2020 Long Range Plan), in order to participate in SKA construction and operations at a level commensurate with Canada's scientific and technological ambitions, are specified in Figure 3. These resources represent 6% of estimated global SKA processing, online and nearline requirements, as determined from the bottom up using anticipated SKA use cases (Bolton+ 18; Quinn+ 20; AENEAS+ 20), in addition to a network capacity of 100 Gb/s to enable data flow between telescopes and SRCs. The timeline for growth in this DRI capacity as well as the functionality (i.e. personnel) to exploit it, from which the integrated numbers in Figure 3 can be derived, is shown in on the right-hand side of Figure 2.

Canadian SRC: DRI Resource Estimates

Timeline	Processing (Cumulative)	Online Storage (Cumulative)	Nearline Storage (Cumulative)	Data transport
2021 – 2024	1.5 PFLOP-yr	33 PB-yr	23 PB-yr	100 Gb/s
2021 – 2030	9.7 PFLOP-yr	238 PB-yr	654 PB-yr	100 Gb/s

Figure 3: Estimated DRI resources for a Canadian SRC from 2021 – 2024 and from 2021 – 2030. Investments in personnel in excess of 30 FTE-years from 2021 – 2024 and 115 FTE-years from 2021 – 2030 will also be needed to achieve SRC functionality.

Investments in personnel in excess of 30 FTE-years from 2021 – 2024 and 115 FTE-years from 2021 – 2030 will be needed to achieve SRC functionality. Figure 2 shows that Canadian investment in the SRC Alliance must begin in 2021 and ramp up towards full capacity and functionality by 2028.

There is a strong consensus within the Canadian astronomical community (e.g., Kavelaars+ 19; Spekkens+ 19; Barmby, Gaensler+ 20) that Canada should host an SRC to fulfil the scientific data processing, storage and user support obligations to the SKA highlighted in Figure 3, as opposed to a model in which those services are purchased elsewhere. Canada's SRC would be the only one within the global SRC network in North or South America, providing the only direct connection for SKA users on our continent. Scientific computing platform and archive development is a Canadian strength. The NRC-operated Canadian Astronomy Data Centre (CADC), the Canadian Advanced Network for Astronomical Research (CANFAR), and the scientific community are well-positioned to collaborate in playing a leading role in the development of SRC architectures and implementations. A Canadian SRC would leverage this leadership to deliver computing, storage, and user support tailored to the needs and ambitions of the Canadian community. Since the funds required to operate a Canadian SRC would be spent in Canada and would include significant investments in personnel who would innovate in the digital landscape, hosting a Canadian SRC would also provide significant DRI return on our SKA investment, much more so than outsourcing that capacity.

Canadian astronomers, the Compute Canada Federation (CCF), the CADC, and CANARIE have worked together to establish real-world estimates for the costing of computing hardware, networking, and personnel that will be required to operate an SRC (Rosolowsky+ 19). The costs are derived from on-going Canadian infrastructure roll-outs and reflect the total cost of system ownership, and are among the most detailed and sophisticated estimates available across the global SKA project. They account for processing, online and nearline storage costs, and networking, in compliance with SRC network guidelines (Bolton+ 18, AENEAS+ 20); 5-year processing refresh costs and continuous storage refresh costs; as well as staffing costs to provide user support for observers, archive users, EPO and EDI initiatives. A processing ramp-up to full operations in the next decade is also embedded in the estimates, with processing hardware purchased and science commissioning processing/storage beginning in 2022, and the capacity to support the science compute load of full operations in 2028. This exercise implies that the total inflation-adjusted cost of a Canadian SRC from 2021 – 2030 is $\sim$ \$90M CAD in then-year dollars, with the bulk of the funds allocated to storage and staffing, with processing and networking costs contributing the rest. Since the SRC is a vital component of Canadian participation in the SKA, support for the associated costs will be requested as part of the broader request to Government for SKA funding in the next decade. Almost all of these funds would be invested nationally, and represents about a third of the total cost of Canadian participation in the SKA through 2030.

Beyond a DRI facility, the vision for a Canadian SRC is that of a hub of SKA activity that fosters digital innovation and has positive broader societal impacts. From a research standpoint, data from NASA flagship missions shows that not only do effective archives increase the scientific productivity of the community (White+ 09), but they also increase scientific accessibility (Peak+ 19), expanding the user base to include a broader national and international community. Partnerships between SRC science support staff and university researchers could effectively leveraging resources in both arenas to boost DRI innovation. The Canadian SKA Regional Centre is also the natural host organization for SKA-related EPO and EDI efforts that are aligned with the recommendations regarding astronomy and society in the 2020 Long Range Plan. Participation in the SKA would mean a major public investment in radio astronomy, and it is critical to make the effort to communicate the return on investment back to the public as well as for societal benefits to be part of that return.

3 NDRIO and a Canadian SRC

Working with the Canadian Astronomy Data Centre (CADC) and Canadian universities, NDRIO will be an essential partner in operating a successful Canadian SRC. The needs of such an SRC reach across all of NDRIO's mandate, including Advanced Research Computing, Research Data Management, and Research Software. There is already excellent precedent for this wide-reaching partnership. The CADC currently operates an integrated science archive that includes the Canadian Advanced Network for Astronomical Research (CANFAR) science platform, developed in partnership with Canadian universities and deployed in the Compute Canada Federation (CCF) cloud service layer. The Canadian SRC plan is based on expanding such a collaboration with NDRIO to scale up this model. This is similar to the approach adopted in particle physics through the CERN/ATLAS project, which is also deployed in the CCF. In the particle physics nomenclature, the Canadian SRC will likely lie somewhere in between a "Tier-1" and "Tier-2" facility. The SKA is one of several large data projects in Canadian astronomy coming online soon (e.g., Canadian Hydrogen Observatory and Radio-transient Detector [CHORD] and the Rubin Observatory [formerly LSST]; see Barmby, Gaensler+ 20 and Lovekin+ 20). The capabilities of an SRC are therefore part of the broader DRI in astronomy that will be required in the coming decade (Kavelaars+ 19).

Some of the capabilities that will be needed for a Canadian SRC are being developed by the Canadian Initiative for Radio Astronomy Data Analysis (CIRADA; https://cirada.ca). CIRADA is funded by a $10.3M CAD CFI grant to Canadian universities and is a parternship between CCF, the CADC, CANFAR, and international research projects. The CIRADA project is developing research software to generate science-ready data products for on-going large surveys on radio telescopes. Some of these processes will run in the CANFAR science platform deployed on CCF services and resources. The generated data products will be made available to the international research community through a combination of CANFAR and CADC services. When completed in 2023, CIRADA will have demonstrated the scientific potential of the SRC model, which comprises the infrastructure, computing capability, and expertise needed to convert the enormous raw data streams from next-generation telescopes into sophisticated digital products that scientists can use to make new discoveries.

> We advocate for a partnership between NDRIO, CANARIE, the CADC, CANFAR and universities to operate a Canadian SKA Regional Centre as a hub of SKA digital innovation that integrates seamlessly and synergistically into the broader national DRI ecosystem.

References:

AENEAS+ 20, Project reports from the Advanced European Network of E-infrastructures for Astronomy with the SKA • Barmby, Gaensler+ 20, The 2020 Long Range Plan for Astronomy and Astrophysics (LRP2020) • Bolton+ 18, SKA Regional Centre Requirements, SKA Organisation Technical Memo • Dewdney+ 16, SKA1 System Baseline Design, SKA Organisation Technical Memo • Kavelaars+ 19, LRP2020 white paper: Digital Research Infrastructure in Astronomy • Kavelaars+ 20, NDRIO white paper: The Canadian Astronomy Data Centre • Lovekin +20, NDRIO white paper: Digital Research Infrastructure for Canadian Astronomy • Peek+ 19, Robust Archives Maximize Scientific Accessibility, American Astronomical Society Meeting #233, id.207.09 • Quinn+ 20, A White Paper by the SKA Regional Centre Steering Committee • Rosolowsky+ 19, Costing a Canadian SRC • Spekkens+ 19, LRP2020 white paper: Canada and the SKA from 2020-2030 • White+ 09, ASTRO2010 white paper: The High Impact of Astronomical Data Archives