Australian SKA Regional Centre Design Study Program

Program Plan (2019-2022)

Document Status	Final
Released	30 April 2019
Developed by	Slava Kitaeff
Approved by	Peter Quinn

1. Reference documents

The following documents are referenced in this document.

[RD1] Quinn et al https://arxiv.org/abs/1501.05367
[RD2] SKA-TEL-SKO-0000706 SKA Regional Centres: Background and Framework
[RD3] AusSRC White Paper, https://tinyurl.com/ybglblml
[RD4] 1st SKA Big Data Workshop report, https://tinyurl.com/yboznhux
[RD5] 2nd SKA Big Data Workshop report, https://tinyurl.com/y73uv2lt
[RD6] R.Bolton, SKA & SRCCG update, AENEAS'18 all hands meeting, 2018
[RD7] M.Wise, AENEAS Meeting Overview, AENEAS'18 all hands meeting, 2018
[RD8] AusSRC visualisation workshop, 5 Nov 2018, https://confluence.csiro.au/x/bAUVLQ
[RD9] SKA-TEL-SDP-0000038 SKA1 SDP System Sizing, 2018
[RD10] Cornwell, T., 2013, Data Products for SKA and SKA Low, http://www.skatelescope.org/wp-content/uploads/2013/04/Cornwell-SKA_Lowscienceassessmentdataproducts.pdf
[RD11] SKA-TEL-SKO-0000735 SKA Regional Centre Requirements
[RD12] SKA-TEL-SDP-0000180 SKA1 SDP High Level Overview
[RD13] R.Bolton, Square Kilometre Array: Current Status, AENEAS, March 2019, https://indico.astron.nl/getFile.py/access?contribId=9&resId=0&materialId=slides&confId=198
[RD14] PMBOK Guide, 2008, Fourth edition, ISBN: 978-933890-51-7

2. Background

Australia will host SKA1-Low, a facility that will produce around 300 PB per year of data products that need to be disseminated to science teams across the globe. SKA project will not have the capital resources to design, build, deliver, and operate end-to-end support for science data products, archives, and associated services. The gap is large – comparable to the current construction and operations budget [RD1].

The SKAO Data Flow Advisory Panel recognized the existence of this shortfall and recommended that the SKAO Board encourage SKA member states to form a collaborative network of SKA Regional Centres (SRCs) to provide the essential functions that are not presently provided within the scope of the SKA1 project.

It is proposed that the SRCs form a Regional Centre Alliance (RCA). The following key factors are discussed in "SKA Regional Centres: Background and Framework" and will need to be addressed by the Alliance [RD2]:

- **Data products**. The generation of advanced data products specific to the science goals is not within the scope of the current SKA Observatory. The combination of data products from across multiple observations must be undertaken outside of the Observatory boundaries by the science teams;

- **Data volumes**. The data volume that will be generated by the SKA will be very high. The data products need to be curated and served to the community while abiding by the SKA's data access policies. It is expected that SKA users will run several models for their analyses, and final data volumes associated with each experiment may also be large. In addition, future upgrades could increase the delivery from each site, or if additional high level science data products are generated. Given the overall volume of data products and the large potential size of each one, a traditional model where users download their data to a local machine to process or analyse is not viable. Instead, SKA users will need to take the processing to the data, submitting requests to run analysis or visualisation workflows on data held remotely;
- **Collaboration.** In the first years of full SKA operations, the science program will be defined by the PI and Key Science Projects (KSPs). The teams are likely draw the membership from across the SKA member nations and the wider international community, conducting ambitious science programs that require thousands of hours of observing time on the SKA. To maximise the scientific productivity of the SKA will require new methods, algorithms, and software. Such activities will necessarily be driven from the SKA scientific community, and will require an environment that enables innovation in research and successful collaboration.

[RD11] outlines six broad areas of SRC requirements:

- **Governance** - How the SRCs are made to function as an alliance and to work serving the needs of the SKAO and the user community;
- **Science Archive** - Hoe users will access the SKA science data products;
- **Storage Capacity** - Provision for the storage, data management and curation;
- **Accessibility and Software Tools** - How users will interact with SRCs;
- **Data Processing Capacity** - Generation and visualisation of science data products;
- **Network connectivity** - Transfer of science data products from the Observatory into SRCs and between SRCs.

[RD12] defines the boundary of SDP and its likely relation to the SRCs as "The SDP allows appropriately privileged external users and users within the Observatory to query the metadata associated with the data products. The required Authentication, Authorisation, Allocation and Identity (AAAI) management information about users is requested from TM. The result of such a query may be the bulk transfer of data to an SKA Regional Centre (SRC) over International WANs and National Research and Education Networks (NRENs). The SKA Regional Centres will be required to support a query client and to support a function to receive data from the SDP. The SDP produces a number of standard data products which are delivered to SKA Regional Centres and also persisted in long term storage as the principal archive of SKA data." (See Figure 1).

The interfaces and the exact mechanism of transferring data from SDP to SRCs are not defined at this stage, and will be a subject of the future work of the SRC design projects and SDP. Some of the planning for the infrastructure and architecture design will depend on the policies and decisions made on how exactly the data will flow from SDP to SRCs, and on the level of redundancy required in data replication.

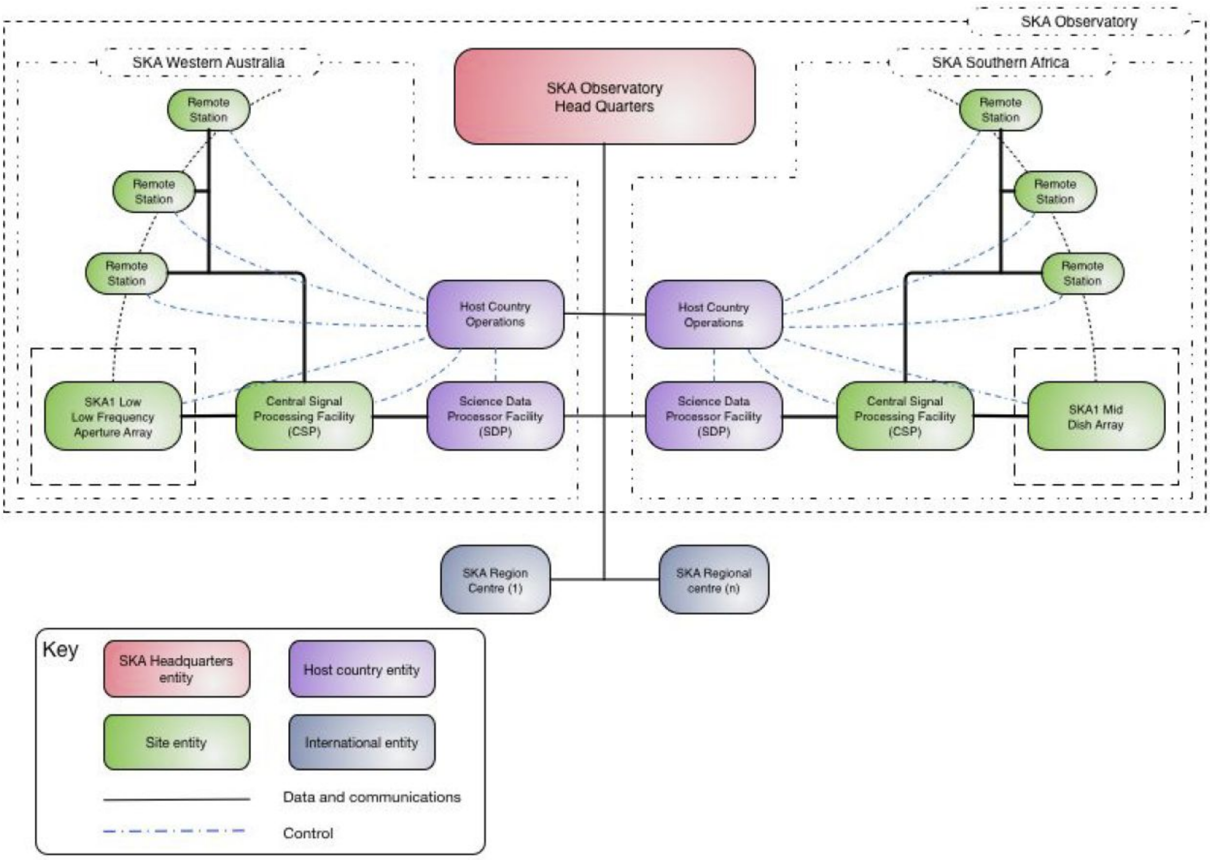

Figure 1. Overview of the SKA systems showing the relationship of the SDP to the rest of the SKA system and SRCs. Adopted from [RD12]

Thus, in short, the scope of SRCs can be defined as:

1. Data flow and data dissemination from the SKA to users;
2. Storing, publishing and curating data;
3. Post-processing and data analyses;
4. Single user portal;
5. User support.

2.1 SRC Steering Committee (SRCSC)

SRCSC is a new global group established to take forward the development of SRCs. It consists of one representative per SKA member country, and is expected to provide much better connections between individual countries and the work that will be done in the next few years to develop SRCs.

Working groups will be established to take forward various aspects of the work. [RD13] provides initial ideas for the focus of the working groups:

- Workload management - compute and data placement;
- Data Lifecycle, Quality of Service requirements;
- Users:
    - Archive Exploration
    - Data interaction and visualization
    - Science Gateway

  - Large job submission
  - Programme users (KSP, PI)
  - User Support
- NREN/ International data transfer;
- SRC operations and resource management.

The program will participate in the established SRC working groups.

2.2 Capacity of SRCs

SDP provides the the Archive size estimate [RD9] as in Table 1. These estimates are relatively modest.

Data	Data rate to Archive / Gbit s^{-1}	Growth rate of Archive / TB $(day)^{-1}$	Growth rate of Archive / PB $(year)^{-1}$	Size of archive after 5 years / PB
SKA1-Mid HPSOs	9	90	34	170
SKA1-Low HPSOs (excluding EoR vis. data)	3	30	11	55
SKA1-Low EoR vis. data	22	220	77	385

Table 1. Archive size estimates. Adopted from [RD9]

However, because of the thought participation of SRC in debugging SDP, the estimate provided by the SKAO is higher. According to [RD13] the SKA Science Archive will grow 600PB per year, once the SKA is fully operational (See Figure 2).

It is possible that SRCs may voluntary accept the visibility data to participate in debugging SDP during construction and commissioning in 2022 Q4 – 2024 (see Table 2). It is also expected that SRC will need to be available prior to the KSP are fully underway. Adopted from [RD6] Table 2 suggests a scenario of development of global capacity of SRCs. It provides some projections on what AusSRC may choose to provide in storage, compute capability and network to the SKA1.

There are multiple scenarios for where data is to be stored. In the maximum scenario, if AusSRC decides to host a full copy of both SKA-low and SKA-mid, the storage capacity by 2024 will need to be at around 300 PB, while the required processing capability might be expected to be at about 70 PFlops[1].

[1] Numbers in previous versions of the document were incorrectly halved in the assumption that the table gives combined numbers for the SKA-low and -mid, while they are actually “per site”.

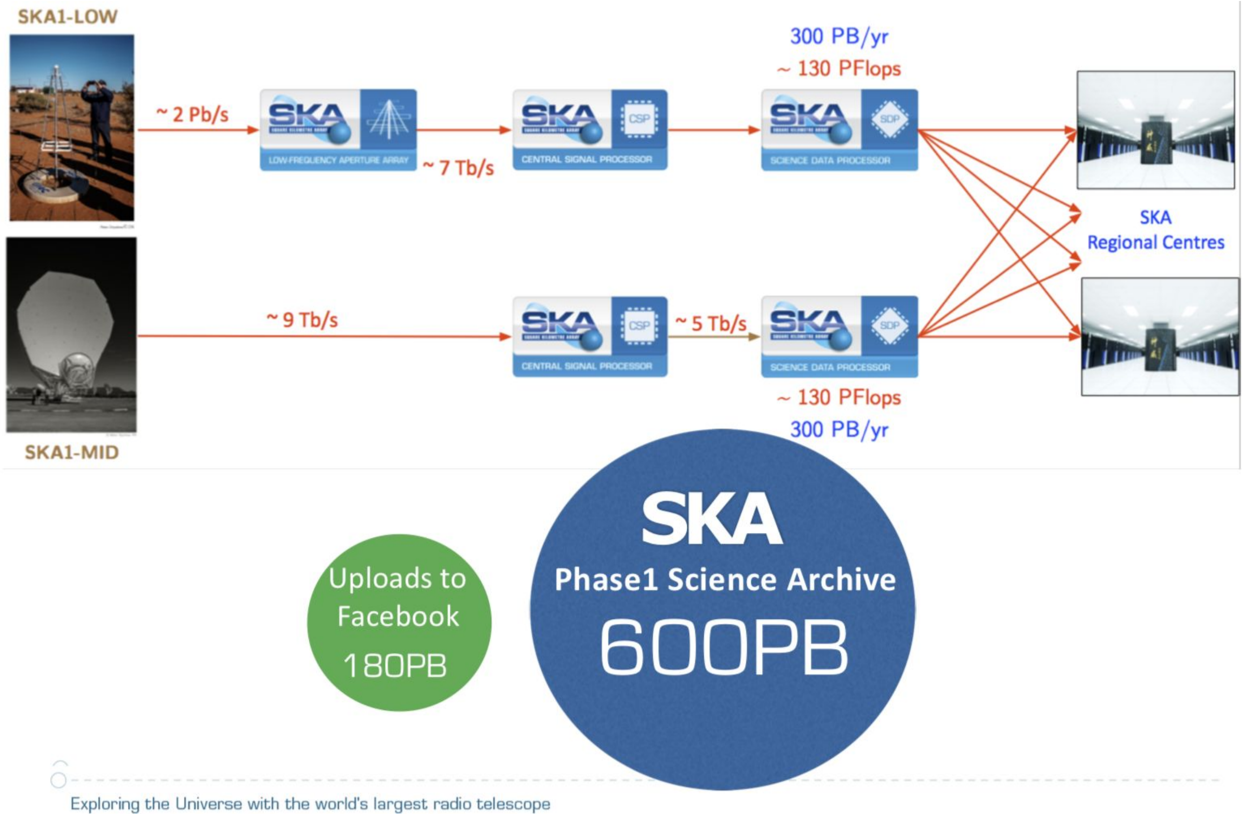

Figure 2. SKA data flow. Adopted from [RD13].

Summary – bare minimum (ideally want headroom)

Year	Array Status	Archive size estimate and processing	Useful data set size	SRC Involvement
Pre-2022	Precursors and pathfinders	5 PB 50 TFLOPS	Up to 0.1 PByte scale	Data challenges with precursor, pathfinder and simulated data.
2022Q4	First dishes / stations on site	5 PBytes 60 TFLOPS	Up to 0.1 PBytes scale for data challenges, 50-200 GByte scale from SKA arrays	Voluntary acceptance of some principally useful visibility data sets, de-bug SDP pipelines and tune parameters. Science Verification work from 2024. Engagement of community experts from outside SKAO Interface testing
2023	Around 10-20% of total	24 PBytes, 7 PFLOPS	Up to 0.1 PBytes scale for data challenges, 1-30 TByte scale from SKA arrays	
2024	Up to around 50-70% of total	**150 PBytes, 35 PFLOPS**	0.02-1 PByte scale	
2025	Building to full scale	**365 PBytes, 35 PFLOPS**	0.1-5 PByte scale	A few Public data products available via SRCs
2026	Complete arrays, shared risk (early) PI science	**900 PBytes, 35 PFLOPS**	0.1-5 PByte scale	All easy modes enabled, commissioning of non-standard techniques Full scale SDP available (but HW may be at deployment baseline level) SRC role takes on project based user support for SRC processing and analysis of observatory products.
2027	PI science	**1.4 EBytes, 35 PFLOPS**	0.1-10 PByte scale	
2028	PI & KSP Science	**1.7 EBytes, 35 PFLOPS**	0.1-10 PByte scale	Full functionality SKA

Exploring the Universe with the world's largest radio telescope

Table 2. SRC global processing and archive capability estimate. Adopted from [RD6].

2.3 SRC in Australia

The National Innovation and Science Agenda (NISA)[2] budget did not include any provision for a proposed SKA Regional Centre as the requirement had not been identified at the time. Any provision for establishing an SKA Regional Centre would be an additional financial implication dealt with in the proposed 2020-21 Budget submission (DIIS Milestone 4 deliverable).
An Australian SKA Regional Centre will be pivotal to the Australian science teams in realisation of science with SKA data, development of a human capital, and Big Data capability outcomes associated with hosting the SKA. The Australian SRC would be part of a global network providing the computational/data intensive infrastructure and support services essential to enabling SKA's science programs and long term data curation and archiving. This network would work closely with the SKA Observatory but would be separately governed and funded.

A significant body of work is required to develop the SRC concept. No similar facility currently exists in Australia (though similar concepts are emerging and being implemented in other contexts around the world). A key strand of this program is for a design study that would identify, assess, and test potential solutions to providing the tools, services, and people needed for the Australian (and regional partner) research communities to utilise the SKA Observatory data outputs. The solution could bring together a range of resources, potentially including Australia's HPC network (including the Pawsey Centre); other HPC capabilities (for example the Swinburne University OzStar facility); commercial cloud services; AARNet; and overseas facilities. The design study would identify solutions and test how effectively they work (i.e. how effectively they manage the workflow processing SKA data required). A particular focus will be on testing scalability – that is the ability of the workflow options to operate at the SKA scale.

During the last two years ICRAR and CSIRO have undertaken a series of preparatory activities.

- In April 2017 the ERIDANUS Project was launched to prototype the elements of SRC infrastructure between and within Australia and China[3].
- In November 2017 the first AusSRC community workshop developed a white paper containing the high level requirements for the SRC in Australia [RD3].
- In March-April 2018 two information sessions (in Perth and Sydney) were organised to inform the Australian industry sector about the opportunities and challenges of AusSRC[4].
- In April 2017 and April 2018 two practical SKA Big Data workshops were conducted in Shanghai [RD4], [RD5].
- In November 2018 AusSRC workshop in visualisation was conducted in collaboration with Pawsey and Data61.

2.4 Program Key Activities

The program is planning the following key activities:

[2] https://www.industry.gov.au/strategies-for-the-future/boosting-innovation-and-science
[3] http://eridanus.org.au
[4] https://eridanus.net.au/?p=182

- Develop AusSRC requirements based on the inputs from the SKA, international SRC development, and Australian SKA Precursors;
- Develop a technical architecture of the AusSRC, prototype its key elements, and develop an implementation plan;
- Execute a range of projects focused on supporting and improving the science outputs of ASKAP and MWA in order to enable the development of precursors driven requirements, and assist the precursors with their specific needs;
- Participate in development and verification of requirements for the international SRC, particular through prototyping and deploying the elements of infrastructure between AusSRC and ChinaSRC;
- Participate in the SKA and SRC data challenges;
- Develop and prototype the architecture of the SKA Science Archive in AusSRC and its API for post-processing;
- Prototype solutions for the visual data analysis of SKA-scale data demonstrated in ASKAP and MWA science projects;
- Develop and present the Australian Government with the AusSRC Business Case/Plan;
- Establish links and collaborations with the international effort to develop the SKA Regional Centres, participate in established SRCSC working groups;
- Establish links, develop strategies and collaborative projects with Australian ICT companies that can assist with achieving the above goals;
- Establish links to other relevant projects such as AENEAS, OzGrav, ASVO, WLCG (computing for LHC) etc.

3 Program Partners and Funding Contributions

#	Source of funding	Amount
1.	DIIS	$2,000,000
2.	CSIRO	$1,800,000
4.	Pawsey	$436,184.00

#	Partners	Interest
1.	ICRAR (UWA)	SRC architecture, science archive, science surveys
2.	CSIRO	ASKAP science surveys, science archive
3.	ICRAR (Curtin)	MWA science surveys
4.	Pawsey	SRC infrastructure, architecture and visualisation

4 SKA and International SRC effort

At this stage most of international effort has gone into identifying the scope of SRCs, and understanding the effort required to enable science with SKA data in SRCs. It is expected that in 2019 the international SRC effort will mature and will obtain more formal status and organisation. The ERIDANUS project has already connected to other SRC projects (see Figure 3) [RD7]. This program will participate in developing a common SRC strategy and work packages with other SRC projects.

The program will participate in the relevant SKA and SRC data challenges initiated by the SKAO and SRCSC.

5 Australian SKA Precursors

While the key objectives of the program are focused on enabling science with the SKA data in Australia, the proposed approach is to use the most challenging (from the point of view of data and computing) ASKAP and MWA science cases to inform the AusSRC design. This approach is also targeted at creating tangible outcomes for ASKAP and MWA through addressing the challenges in post-processing, storing, visualising, and analysing the data. Participating in ASKAP and MWA projects will allow the program team to develop transferable skills and expertise for the future SRC.

It is envisioned that the program personnel employed by the partner organisations will be deployed by the program within ASKAP and MWA teams to assist with specific projects to create tangible outcomes for ASKAP and MWA, and to discover in-depth requirements for the AusSRC architecture. The Program Lead (PL) deploys the personnel on tasks as the project/task skill-set required.

Figure 3. SRC - Strategic Partnership. Adopted from [RD7].

5.1 Project proposals

The PL plans and manages the program resources, and makes a call for proposals, whenever the resources may become available for new projects. The project proposals are submitted to PL, reviewed and presented by the PL to the AusSRC MC for approval.

The project proposals will demonstrate:

1. The proposed project leads to at least one of the SKA KSPs, and the proposed tasks are within the scope of SRCs;
2. The proposed project helps to collect in-depth requirements for the AusSRC;
3. The project will help to address data and computational SRC challenges;
4. Commitment of the science team to the proposed project.

It is envisioned that most of the precursor projects will be in post-processing and data analysis optimisation, data architecture, and visually aided data analysis.

6 Community consultations and Asia-Pacific Regional Effort

The ERIDANUS project will continue as an Australia-China regional effort to prototype and deploy the elements of SRC infrastructure, now as part of the program. Annual Australia-China workshops will continue with the next one planned for August-September 2019. While the outcomes of this part of work are not aimed directly at any operational aspect of ASKAP or MWA, some of the outcomes will have the potential to be tangible for the precursors. The decision to use in production any prototype or solution produced by the program is made by ASKAP and MWA. Any assistance required for deploying the solutions in production needs to be negotiated and decided in consultation with the program.

The program will spend $135,000 over the next 3 years on organising national and Asia-Pacific workshops, hackathons, and meetings to continue developing specific requirements, executing trials of new software and infrastructure, assisting the science teams with solving data and computational challenges, and developing a broad range of expertise of the program's personnel.

7 Industry engagement

The Australian Government has requested the program to provide advice on how the Big Data industry can leverage development of an AusSRC. Consultations with the Digital Industries sector begun in March 2018, and strategies on engagement are being developed in collaboration with ResearchBiz, AtomicSky, and CISCO Innovation Central. ResearchBiz has been contracted [5] to produce recommendations on channels of engagement with digital industries for innovation, and a matrix of Australian digital capabilities mapped onto the high level requirements of AusSRC as outlined in the White Paper. The report is expected to be available by the end of March 2019.

[5] Previous AusSRC Industry Engagement UWA contract, not part of this project budget.

The program allocates $215,000 over the next two years to engage with industry in specific projects to test some of the recommended channels of engagement. Such projects should cover areas of expertise that are not available within the program.

8 Personnel Training

We aim that the program personnel will develop expertise and becomes the core team of the future AusSRC. This team needs to acquire broad expertise to become a pool of experts that can be deployed on different tasks and projects. The program will run multiple tasks and projects contributing to the development of a holistic vision of the future AusSRC. Running many development projects, and providing the maintenance and user support at the same time requires a systematic approach. This is not a new concept, and industry has developed a range of methodologies and tools to achieve the agility, continuity and robustness of processes.

One such widely adopted methodology is DevOps, which is a combination of software development (*Dev*) and information technology operations (*Ops*). It is a set of software development practices that aim to shorten the systems development life cycle while delivering features, fixes, and updates frequently in close alignment with business objectives[6].

Another method of distributed agile development, SAFE, has been adopted by SDP SKA, and ICRAR DIA and SC CASS teams had been already certified for SAFE.

The program allocates $32,450 to conduct training and certification of personnel, mostly in the first year of the program.

9 Infrastructure

By 2023 SRCs are expected to have a combined compute capability of 7PFlops and 5PB of storage capacity, ramping the capability to 150 PFlops and capacity to 35PB in the following year [RD6]. Such an infrastructure must satisfy a whole spectrum of data-intensive requirements and be suitable for delivery of data-as-service, software-as-service, and infrastructure-as-service solutions. To provide maximum efficiency, such infrastructure needs to be hardware/software co-designed. Working with Pawsey, ICRAR, Data61, CSIRO, SKA precursors, other SRC projects, and industry, the program will investigate and prototype solutions, and undertake SKA data challenges on prototyped architectures.

Pawsey and the program will work together on studying and testing technical solutions into the future, as well as developing a model of co-maintenance and user support for the future AusSRC. The program will appoint a System Specialist to lead this part of work. This position will help to transfer requirements, test solutions, commission and operate the system, and participate in the development of the architecture and prototyping activities.

The program also allocates $130,000 to be spent on software and cloud services. Purchased cloud resources will be partly used for development and prototyping, and partly will be made

[6] https://en.wikipedia.org/wiki/DevOps

available to ASKAP and MWA research teams involved in prototyping activities to process data.

10 Program Governance and Management Model

It is important than the program takes a holistic approach to developing an understanding of how the future AusSRC should enable science with SKA data. The program governance and management model will be critical to the success of the project, especially due to its distributed employment model.

10.1 Governance

The future structure and governance of an AusSRC, as part of the AusSRC Business case, will be defined by the AusSRC MC during the course of the program.
With regard to the program:

- The Program Lead:
    - reports to the MC for all matters related to the delivery of the program agreed outcomes;
    - works with the precursor projects and Pawsey management to define a set of projects that will be executed as part of the program and presents them to the MC for approval;
    - staff employed with the program funds will be line managed through their host organizations and will work within the program team with responsibility for the program deliverables to the PL.
- AusSRC MC will ensure collaboration with international SRC efforts to align the program as required;
- AusSRC will ensure communications and alignment with other national and regional efforts relevant to the AusSRC;
- The AusSRC MC reports to ANZSCC on the program outcomes.

10.2 Management

To achieve the program's objectives several management principals will be used:

1. The program personnel is employed through partner institutions.
2. The program personnel reports to the PL, who deploys the personnel on the program tasks and projects.
3. The PL plans the projects with the external science and technical teams identifying the project objectives, requirements, outcomes/deliverables, commitments, milestones and resources;
4. The program will adopt best software development practices as applicable to the specifics of each project and the program as a whole. The program will train the personnel to achieve high efficiency and productivity;
5. The program and project management will include:
    a. Weekly program meetings;
    b. Weekly or fortnightly project meetings, as applicable;
    c. Common work space;
    d. Common software tools (JIRA, Confluence, Slack, etc);
6. The host institutions will provide:

a. The office space for the employed program members;
b. Agreed numbers of floating desks;
c. Assistance with introduction and integration of the program personnel with science and/or technical teams.

11 Technical Domains

The SRCs will have to address a number of serious issues that lead to increasing cost of the required infrastructure, inefficient use of resources, and as a result, low scientific output from large investments into the infrastructure and instruments. Specifically:

- inefficient software written by students and untrained scientists to post-process and analyse data from the telescopes;
- use of generic platforms that are not optimised for data intensive applications;
- disconnected views on the subsystems that enable scientific discovery in data.

The program will take an approach that will address these shortcomings through a holistic view of what SRCs will have to provide to enable science with the SKA and its precursors. The main goal is to make SKA-scale data post-processing efficient and effortless for the science user.

Figure 5 depicts technical domains that needs to be developed for the future AusSRC.

The three system domains - Science Archive, Post-processing, and Visualisation will inform the System Architecture that enables interactions between the components via the interfaces and communication protocols.

The System Architecture will drive the requirements on the Infrastructure domain that also needs to be prototyped with a view of forthcoming and prospective technologies.

The Applications and Algorithms domain will use the external interfaces and protocols to interact with the system components to do specific user defined tasks.

Finally, the Operations and User Support domain will define how AusSRC systems are to be maintained and operated, and what support users need to be provided.

11.1 System Architecture

It's important to have a proven design of the system architecture of AusSRC as it will be a significant investment. The program will collect in depth requirements working with ASKAP and MWA, as well as SKA1 and SRCs requirements . Such requirements will drive the design of the System Architecture.

Based on previous experience, we know that classical HPC systems such as Magnus are not optimal for the data intensive type of processing required in radio astronomy. The program will work in close collaboration with Pawsey assisting with developing the specifications for astronomy systems at Pawsey.

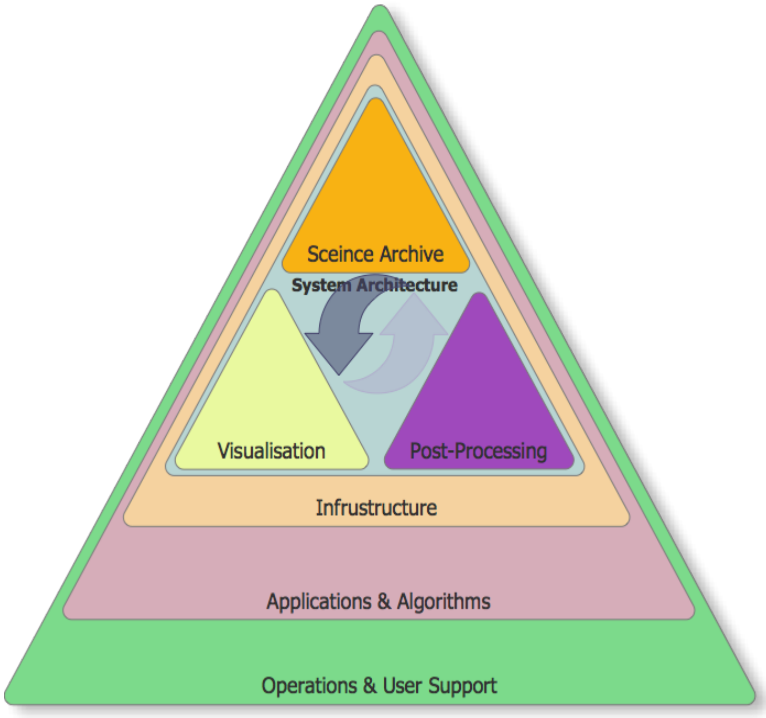

Figure 5. System Domains

This project could also study and model a hybrid HPC-cloud architecture. Part of the effort may be directed at developing sustainable budgeting/accounting models and possibly tools for the hybrid architecture. The project will work with Pawsey on development of the AusSRC system architecture and its prototype to demonstrable an integration with at least one commercial cloud.

11.2 Science Archive

Both ASKAP and MWA have developed Virtual Observatory (VO) compliant portals to the science archives in Pawsey, CASDA, and MWA-ASVO, respectively. These two systems use different implementations of VO interfaces and mostly exploit the "download" framework. Although, such a framework will remain to be useful for data discovery and downloading small datasets, it may not meet the SRC requirement to minimise the data movement and data replication. To satisfy such a requirement we may need to design an in-archive processing and visualisation architecture[7].

Currently, VO does not provide a standard API to access the data in archive directly from an HPC cluster, however the need for such an interface has been discussed and recognised by IVOA.

The program will design, prototype, and propose for adaptation by the IVOA and international SRC such interfaces. The current ICRAR DIA PhD project on optimisation of data warehouse

---

[7] There's a common notion in discussions of processing SKA data that the algorithms need to be "brought close to data", because the data is to big to be moved. However, this notion is insufficient as it does not reflect the need to have a system architecture design that minimizes the **need** to move or replicate the data for processing, inspection or analysis purposes. This is why we use the term "in-archive" to emphasise the importance of the holistic view on the architecture design of SRC, including the back-end of the archive and the interfaces into the archive other than VO download interfaces.

and data analytics platforms, such as Apache Spark, and databases such as Hive will also help to explore the options for an efficient and future proofed solutions.

Some of the designed and prototypred solutions could be considered for adaptation by CASDA or MWA-ASVO. Such adaptation or implementation of the production code are not part of the program objectives directly, however the program may act in a consulting capacity if requested.

The program will collaborate with other SRC projects to have a joint effort in developing the solution. The program will work with ASVO on any relevant issues.

11.3 Post-Processing

The efficiency of post-processing software (mostly Level 6 and 7, as defined by [RD10] (see Figure 6) but in some cases lower levels of data products) on a specific platform, as well as the suitability of a platform for data intensive processing, will be a focus of the program. This task is specifically challenging because the requirements for different types of data and algorithms can significantly differ. Knowing these requirements is critical for the development of robust and efficient system architecture.

At the same time the data processing/post-processing software needs to be optimised for the architecture. The specific challenge is that the pipelines are very different, and these variations need to be studied and understood.

Through assistance in improving the efficiency and performance of the ASKAP and MWA post-processing and data analysis software, and adopting the software to new and more efficient execution frameworks we will develop requirements for the system architecture, science archive, visualisation, and the overall infrastructure of the AusSRC.

The program will collaborate with ASKAP, MWA and the Rialto project (ICRAR DIA and CASS SC) in developing new efficient software solutions, HPC and cloud computing platforms.

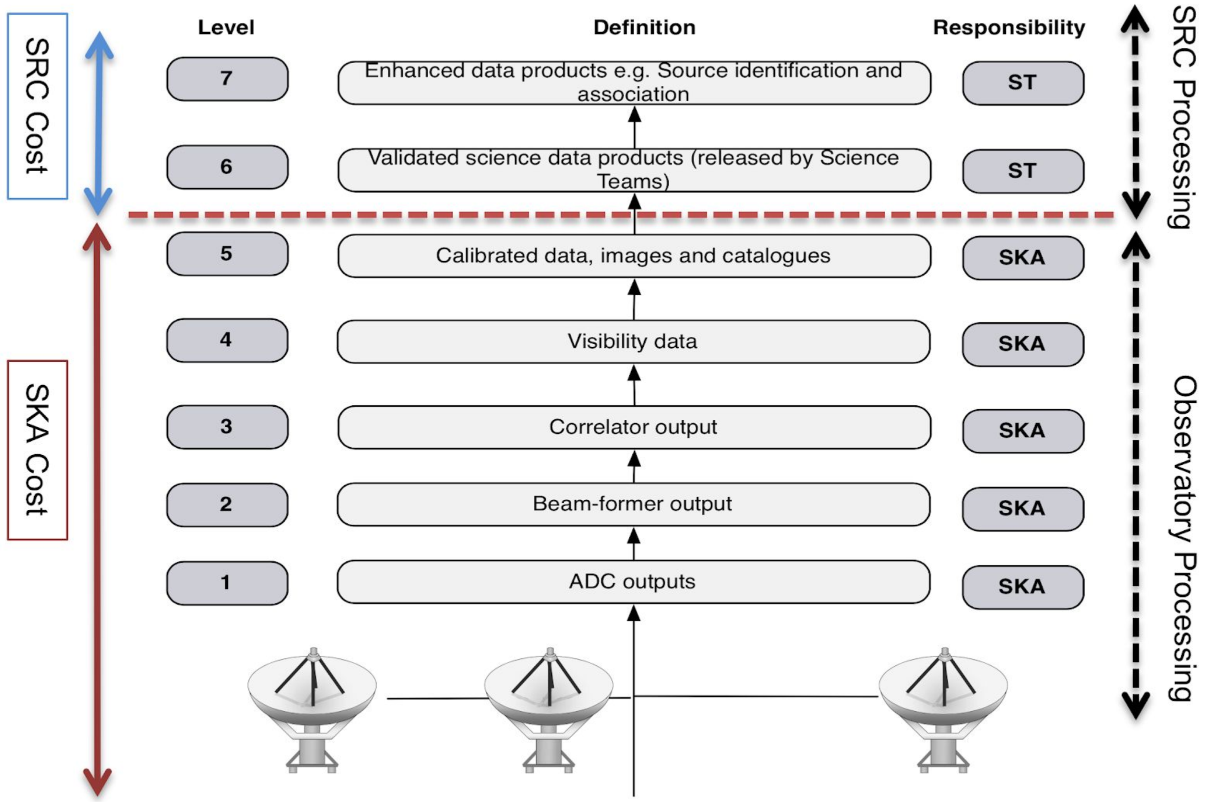

Figure 6. SKA data levels (based on [RD10] classification) and processing.

11.4 Visual Data Analysis and Data Quality Assessment

During consultations, visual data analysis and visual data quality assessment have been identified as areas where no satisfactory software and infrastructure solutions exist, even for precursor data sizes [RD3].

To understand the issue further, a workshop on radio astronomy visualisation involving MWA and ASKAP scientists, DIA and CSIRO Data 61 Visual Data Analysis was held last year [RD8]. Several science use-cases for visualisation were presented and discussed, including spectral-line cubes, visibilities, all-sky, pulsars, 2D, projected 3D, 3D etc. It has been shown that any solution for the SKA-size datasets will need to be remote, distributed, and in-archive.

Pawsey will appoint a Visualisation Specialist to work with the program developing the AusSRC architecture and infrastructure. The main goal will be to design an in-archive remote visualisation solution that will initially be tested and used by precursor survey scientists, and will inform SKA-ready AusSRC architecture.

The program will also sub-contract Data61 Visual Analytics to adopt original visualisation solutions to radio astronomy use-cases.

11.5 Applications & Algorithms / Operations & User Support

The AusSRC project team will need to develop skills and expertise in assisting the science groups with development of algorithms, software, provide the platforms and frameworks for development of applications, training, user support, and develop a general understanding of how to operate an SRC in coordination with other SRCs nationally, globally and the SKA Observatory.

Appendix A. Work Breakdown Structure

Appendix B. Budget

Appendix C. Projected Effort