Australian SKA Regional Centre Workshop

(27-29 Nov 2017, Perth)

Summary of Discussions

Table of Contents

List of abbreviations

ARSC ………………….. Australian SKA Regional Centre
CSIRO ………………… The Commonwealth Scientific and Industrial Research Organisation
ICRAR …………………. International Centre for Radio Astronomy Research
KSP…………………….. Key Science Projects
RCA .... Regional Centre Alliance
SDP .... Science Data Processor
SKA .... Square Kilometre Array
SKAO .... SKA Observatory
SRC .... SKA Regional Centre
SRCCG .... SKA Regional Centres Coordination Group
VO .... Virtual Observatory

1. Background

On 27-29 November 2017, in Perth, ICRAR and CSIRO hosted an Australian SKA Regional Centre Workshop. The workshop was the first broad discussion of the ASRC with the participation of SKAO (Antonio Chrysostomou, UK), LHC (Ian Bird, Switzerland), FAST (Youling Yue, NAOC-China) and China SRC (Tao AN, SHAO-China). 39 workshop participants represented different research groups across Australian institutions involved in SKA and SKA precursors.

The workshop focused on:

- Overviewing the current activities, science and technical projects that are within the scope of SRC.
- Developing the high level requirements and a roadmap for Australian SRC, with an initial focus on data from the Australian precursors (MWA and ASKAP).
- Discussing collaboration models and common goals of Australian SRC with other SRCs around the globe, especially in the Asia-Pacific region.
- Learning from the international regional science centres of other large science projects around the world.

The first day of the workshop was an overview of SRC related developments in Australia and at the international level. The presentations can be downloaded from the ERIDANUS website[1].

The rest of the workshop was structured as a series of focus group discussions. 17 questions were proposed for discussion broadly covering: SKA1 science topics of interest to Australian groups, precursor science and experience of development; needs and expectations of science groups from the future ASRC and global Alliance of SRCs. Three to

[1] https://eridanus.net.au/?p=170

five discussions were held at the same time, resulting in total of 50 discussions over 2.5 days. The following chapter summarises the discussions.

2. Summary of discussions

Some of the discussions were more productive than others due to ability of people to attend and contribute. This summary includes only the discussions that have produced a substantial output. The content will be updated with more contributions at later time.

Common discussions

Multi-messenger / multi-wavelength science in ASRC

Contributors: Simon Driver, Ray Norris, Ryan Shannon, Antonio Chrysostomou, Minh Huynh, Martin Bell

Internal (SKA) triggers may initiate follow-up on a variety of international facilities, which potentially will need to be processed by, and archived through, the Regional Centres.

Internal (SKA) Triggers may also want to compare new data to prior (image-plane) stacked data, generated and held in the Regional Centres, and to external datasets - either over the VO or possibly through repositories held by the Regional Centres (e.g., LSST). This may require MoUs between either the SKA office/regional centres and external organisations. RCs could honour proprietary periods where necessary, while allowing early access for some purposes defined by the MoUs.

External (e.g., GW, GRB) triggers will elicit SKA data which will eventually filter into the Regional Centres, and need to be labelled as such (housekeeping), and paired with the trigger data/event.

Photometric and spectroscopic redshifts

While photo-z's may exist, more optimal photo-z's may arise from the combination of multiple external datasets, e.g., LSST+Euclid+WISE. Potentially LSST+Euclid+WISE will collaborate and hopefully(?) provide photo-z's etc. It may also be desirable to have optimised photo-z's for specific purposes/classes (e.g., AGN), requiring in-house processing of optical/near-IR data. Optimal photo-z's require forced photometry to generate consistent flux measurements with consistent errors suitable for SED or photo-z analysis. One can move LSST+Euclid to SKA, extremely unlikely to move SKA to LSST+Euclid. This presents a unique (science) opportunity for SKA to be the data federation king, facilitating science only possible from a joint SKA+LSST+Euclid archive (interlinked data lakes: lagoons).

Priors and cross-calibration

Use, and easy access and availability, of optical data will be critical in identifying radio source class, e.g., two SF-galaxies v AGN with lobes. Similarly, the prior knowledge of an optical counterpart may enable source detection to lower sigma levels than without potentially doubling, tripling radio catalogue sizes. The prior argument points strongly to the need for the RCs to host external datasets.

Level 3 products

Discussion concluded having raised the issue of Level 3 (data derived) products, e.g., location of known groups, filaments, voids, optical inclinations/sizes, stellar masses, star-formation rates, known redshifts, metallicities, ages etc. Many of these are derived from analysis using federated data, either the RCs can download Level 3 products and make available, or provide them internally. The latter would require optical/near-IR experts embedded in the RCs.

Users publishing data from SKA precursors and SKA1 in ASRC

Contributors: Markus Dolensky, Andreas Wicenec, Tobias Westmeier

SRCs will need to take care of data publishing, which is a case of long term preservation. Such publishing will need to have a number of attributes:

- Linking publications to data products should be supported thereby providing means for assessing the quality and applicability of data products for a given purpose.
- Version control should be mandatory for all published data.
- SRC should provide a framework and tools for the incremental data release, management and curations.
- Tools to populate data products and catalogues should be available.
- SRC should have the policies around data (and software) publishing.

Collaborative post-processing and analysis in SRCs

Contributors: Slava Kitaeff, Chen Wu, Andreas Wicenec, Markus Dolensky

Collaborative post-processing links the teams and team members across Australia and globally for a range of complex tasks such as e.g. combining interferometer and single dish data or multi-messenger data, improving the pipelines and calibration, etc. Such collaboration needs to happen as in virtual as in physical space. SRC should be able to provide both.

In physical space ASRC should:

- Provide an equipped facility for the workshops and “busy” weeks in Perth and have some virtual presence in Sydney, Melbourne and Canberra. ALMA RC model can be used.
- Provide the experts for those workshops and “busy” weeks.
- Have travel grants to ASRC.

In virtual space ASRC should:

- Provide the tools to support science team collaboration, as well as collaboration between the teams that keeps a track of discussions.
- Provide the data management and access policies and tools.

- Provide the provenance of data and software at any stage.
- Provide the environment for software development.
- Provide the tools to interrogate and inspect the data (visualisation).

ASRC should have a pool of expert that can assist the teams with the algorithms and software development, complex data analysis (e.g. Machine Learning and Statistics). Having such a pool of experts it make sense to share the experts amongst multiple disciplines such as Particle Physics, Gravitational Waves and others. These experts can also be linked to the industry, and as provide the consultancy to the industry as to draw the expertise from the industry if necessary. ASRC should foster the industry connections in general.

SRC archives, portals, interfaces and data access policies

Contributors: Markus Dolensky, Bart Pindor, Katie Jameson, Attila Popping, Phil Edwards, Ray Norris, James Allison, Dane Kleiner

SRC policies should include:

- Prioritization of resource allocation
- Principles of team management
- Responsibility for the derived science data products
- Data and software stewardship
- Define the rules for data releases
- Ensure uniform access across SRCs

SRC portal should provide:

- Uniform access across the Alliance of SRCs
- Federation of identity
- Visualisation capability
- Web interface
- Data discovery capability
- Cross-matching capability across all wavelengths
- Programmatic API
- Tools specific to radio (large FoV, large data sets)
- Should be possible to assign specific user roles
- Tools to manage data releases

Role of ASRC in training a new generation of leading scientists

Contributors: Martin Bell, Chenoa Tremblay, Bi-Qing For, James Allison, Sarah Pearce, Dane Kleiner

SRCs will have a concentration of experts in data intensive astronomy, computer science and SKA scientists. This has the potential to make SRCs an essential part of training of next generation of leading scientists. Some recommendations can be made for the future planning of ASRC to ensure that such opportunity is not lost.

SRC should:

- Practice joint appointments of staff with the universities and joint supervision of postgraduate students.
- Prioritise postgraduate training and co-supervision
- Offer scholarships (summer, workshop, long-term).
- Facilitate conferences focused at postgraduate training.
- Support industrial placements for co-supervised students within SRC in the fields of big data and computer science.
- Provide an ongoing upskilling training to current scientists.
- Integrate with University programs providing the country (world?) wide delivery of courses that are accredited with the universities.
- Allocate the telescope and compute resources to facilitate training.

Role of ASRC in training computing and data skills

Contributors: Ian Bird, Bi-Qing For, Ryan Shannon, Youling Yue, Slava Kitaeff, Martin Meyer, Ron Ekers, Bart Pindor, Chen Wu, Katie Jameson

Training of science students and young scientists in computer science is important for their long term career prospects - both in science and in industry. SRC could be a centre for such training.

Another reason for such a training is the long term maintenance and usability of the software.

There is an expectation of help from the ASRC to scientists to optimise their applications to make best use of HPC and other systems; this help includes both

- Training and
- Consultancy (expertise, advice) on optimisation/adaptation

There was a discussion on what software packages may be provided by the SRC rather than individual science projects.

The SRC should in general have a significant role in ensuring that the applications are optimised, in order to make effective use of resources.

Three types of skills have been identified that are required in ASRC in order to support, maintain, and optimise software:

- Scientists - who have the domain knowledge and can develop the algorithms
- Computer scientists - who can implement the algorithms in efficiently written software, and who can act as consultants to help optimise performance, adapt algorithms to be efficient, etc.
- Software engineers - able to provide the overall software environment for long term maintenance and operation, etc.

It has been recognised that the career outlook of astronomers specialising in software can be a problem within the academia. ASRC can fill this gap via training and employing those astronomers as domain experts with sufficient technical background.

Such training can be done in the form of hand on workshops and online courses that ASRC should facilitate. Specifically, SRCs should hold SKA software bootcamps and schools.

SRCs should provide an environment to bring together scientists and software/computing experts.

SRC should play the key role in mapping the science requirements to SKA postprocessing onto implementation of the infrastructure and software.

As the SRCs will design, implement, deploying the post-processing pipelines, they should facilitate the specific training on how to use the facilities and the software in effective way.

As the SRCs will need to provide the software repository and support for common tools, libraries, modules, templates, they will need to train the users about the effective use of such tools and systems. It can be expected that this will improve the reuse the software components and reduce the fragmentation of solutions.

SRCs should provide essential communications between experts (HPC, science, etc), and a channel to vendors/access to experts. Ongoing training should be of part of such facilitations of communications in which SRCs should play a leading role.

How ASRC could address the growth of the trans-disciplinary research

Contributors: Ron Ekers, Slava Kitaeff, Ray Norris, Ian Bird, Katie Jameson, James Murray, Antonio Chrysostomou

The following disciplines have been identified as having links to SKA science:

- Gravitational Waves
- Particle Physics
- Dark Matter
- SETI
- Astrobiology
- Ionosphere
- Planetary science
- Cosmic Rays
- Cherenkov Telescope Array (CTA)
- Cubic Kilometre Neutrino Telescope (KM3NET)

In addition, the following disciplines have been identified having technology links to the SKA

- Medical Imaging
- Defence
- Space Debris

- Bioinformatics
- Earth Observations

From the above two lists the following disciplines have been identified as having potential interest in collaborating on the development of Regional Science Data Centre:

- Gravitational Waves
- Particle Physics
- Dark Matter
- Defence
- SETI
- Astrobiology
- Ionosphere
- Planetary science
- Cosmic Rays
- Cherenkov Telescope Array (CTA)
- Cubic Kilometre Neutrino Telescope (KM3NET)

In the development of SRC, we should also consider those disciplines that would share common infrastructure with us.

However, the development of SRC should not be done in a competition to other national facilities.

The benefit of multidisciplinary approach in developing ASRC should be seen in knowledge exchange across the disciplines.

ASRC should take the multidisciplinary approach in order to make the funding proposals stronger and more compelling.

The key issues of data management and post-processing to be addressed by SRCs

Contributors: Andreas Wicenec, Ian Bird, Ray Norris, Phil Edwards, Antonio Chrysostomou, Youling Yue, Bart Pindor, Chen Wu, Tao An, Minh Huynh, Randall, Peter Quinn, Ian Bird, Tobias, Ryan, Martin Bell

SRCs should:

- Provide the expertise and best practices to preserve products.
- Develop the policies that utilise FAIR principles.
- Assume the time horizon in preservation and provenance 50 years or until superseded.
- Provide the traceability (DOIs and IVOA identifiers) from ‘raw’ data to paper and backwards.
- Provide a way to reproduce results. (Preservation of the full platform may be need. How far do we go?)

- SRC should have policies and operational models for ingestion of data into the alliance of SRCs (who is keeping what).
- Have a policy on how the resources are allocated to the projects.
- Some SRCs must already exist before KSPs (e.g. EOR) can be scheduled.
- Provide the data transfer monitoring, benchmarking and priority queues.
- Provide a notification system to inform PIs on events.
- Provide the expertise for workflow optimisation.
- Provide the formal ways to put data in preservation.
- Provide versioning of data, pipelines, software etc.
- SRC should have a common API for everything
- Support QC during processing (visualisation), support comparison between different runs.
- Support the provenance of pipeline runs
- Provide a solution for the large scale source finding

The Alliance of SRC should provide the international level of coordination for the resource allocation, provenance, software. Will SRCs need to adopt a tiered hierarchy? Although, the underlying hardware might be of heterogeneous nature, the software must be a lot more homogeneous across the SRCs. How do we coordinate the development of such a software?

Preservation of developments and innovations in algorithms

Contributors: Slava Kitaeff, Chen Wu, Andreas Wicenec, Bart Pindor, Markus Dolensky

Algorithms and software are legacy of SRC. SRC should play a significant role in improving the quality of software, algorithms, and reducing the waste of effort in developing of new software and algorithms in general. In particularly, SRCs should:

- SRC should provide the repositories of software, and should have a mechanism of long term preservation for the software and algorithms.
- SRC should provide training for the best practices in coding.
- Provide a cost/benefit analysis framework for new software, including common metrics.
- Provide a mechanism of software maintenance.
- Provide a model in which organisational, community and individual development could co-exist and benefit from each other.
- Promote a social and collaborative responsibility for the community software.
- Provide and promote the standards for the protocols and APIs (VO in particularly).
- Provide plugin the mechanisms in the system.
- Promote a development and use of common astronomy software (e.g. astropy, ascl).
- Develop recommendations for software licensing.
- Provide a common framework for documenting the software.
- Provide a mechanism for the quality control of the software as part of software publishing process.

During the development of the algorithms and software intellectual property with a potential commercial value might be created. ASRC should develop a commercialisation model and have industry and R&D partners.

Citizens SKA science in ASRC

Contributors: Ivy Wong, Slava Kitaeff, Youling Yue, Ray Norris, Katie Jameson, Ian Bird, Antonio Chrysostomou, Martin Meyer, Minh Huynh, Dane Kleiner

Types of citizen science (CS) that currently exist:

A. Interactive (Galaxy Zoo). This projects are useful for producing the training sets for the development of advanced source classification algorithms, though bias must be observed. This types of projects is up to the science teams to setup.
B. Passive (SkyNet, SETI@Home, Boinc). ASRC may usefully harvest the compute resources for many projects. The resources might be transparently allocated through SRC portal. ASRC should include this development in the plans.

CS can be also useful as an outreach and education tool.
ASRC citizen science component could be used as a powerful teaching, outreach and human resource.

Although SC is not a high priority for ASRC, can significantly increase the visibility of SKA in Australia, and create a positive public perception and a sense of importance of astronomy.

ASRC should consider creating dedicated CS fellowships.

User support that ASRC should provide to Australian SKA community

Contributors: Ron Ekers, Minh, Ray Norris, Tobias, Ian Bird, James, Ryan, Slava Kitaeff, Ivy Wong, Youling Yue, Antonio, Katie Jameson, Bi-Qing, James Murray

ASRC should:

- Facilitate busy weeks / workshops / hackathons to solve practical issues
- Provide the framework for testing software
- Provide training to the users how to use SKA
- Provide training to the users how to use SRC
- Provide training to the students at all levels
- Provide access to the experts on data reduction, software and algorithms, telescope, specific observations (spectral line, continuum, polarization etc)
- Provide an assistance/wisdom/advice in planning the observations with SKA and SRC resource allocation (storage and computing)
- Sufficient resources to enable archiving, post-processing, data analysis
- Provide information and assistance to non-radio astronomers to understand radio astronomy data.
- Provide a curated user archive
- Be a facilitator of SKA outreach
- Develop SKA and astronomy ambassador program

- Provide a line of communications to SKAO
- Provide a line of communications to Australian Government and funding agencies
- Provide the users with a discussion platform (chat rooms, forums, social media)
- Provide physical space and communication equipment for meetings and group work
- Provide funding for traveling to ASRC
- Provide links to hardware software vendors/developers for the optimisation work
- Be a science and technology centre rather just another supercomputing centre
- Facilitate the links to other centres around the world

SKA Precursors

Notes on MWA Phase 3

MWA Phase 3 will be closer to SKA1-mid in data sizes, and it's a large distributed collaboration, which makes MWA a good ASRC test platform for:

- SDP
- Archiving
- Data/resources allocation and access
- User support model
- Data dissemination and distribution
- Cluster of expertise
- Specing the hardware with radio astronomy applications

Notes on ASKAP spectral-line observations with fully operational telescope

Contributors: Tobias Westmeier, Martin Meyer, Bi-Qing For, Kate Jameson, James Allison, Attila Popping, Tao An, Dane Kleiner, Y.Yue

Science

- HI absorption
    - Requires continuum model;
    - Optical priors for spectral extraction;
- HI emission – extragalactic
    - Automated source finding;
    - Storing and accessing source catalogues.

Quality control

- Provision of basic quality control metrics with data;
- Identification of scheduling blocks;
- Extraction of basic metrics:
    - Robust RMS measurement / comparison with expected RMS;
    - Simple image arithmetic;
    - Information on data flagging;
    - Basic calibration information (using simple, global projections).

Simple operations on data

- Parameterisation of sources;
- Creation of moment maps;
- Comparison and overlay of data products;
- VO-like interface (e.g. Aladdin);
- Source finding;
- Extraction of spectra;
- Cross-matching (e.g. with optical data);
- Simple arithmetic operations on data cubes;
- Handling of frequency axis:
  - Conversion between reference frames;
  - Conversion between frequency – redshift – velocity.
- Ability to produce and access full-resolution spectral cubes;
- Ability to extract arbitrary sub-volumes or super-volumes of the data (feedback on expected data size / time requirements);
- Accessing the data at different resolutions;
- Continuum images at different frequencies to facilitate HI science.

Data handling

- Ability to import, store and reprocess raw data from ASKAP;
- Modular, flexible data processing pipeline;
- Flexibility to adapt to changes in processing and analysis resulting from progress in scientific/technical understanding of the data;
- Combine data with data from other telescopes;

Spectral line and image analysis

- Ability of carry out measurements on data, e.g. fitting of models;
- Ability to automatically analyse a large number of sources.

Data visualisation

- Both on-site and remotely;
- Ability to overlay catalogues;
- Comparison of multi-wavelength data.

Expected support from ASRC

- Data reduction support;
- Support for data interpretation for radio astronomers and non-radio astronomers;
- Support and preservation of data reduction and analysis software
- Support for line transitions other than HI, e.g. OH, recombination lines;
- Combination of ASKAP and single-dish data, e.g. joint deconvolution;
- Provision of multiple cubes at different resolutions;
- Allocation of limited resources based on need and with consultation of users.

Notes on ASKAP continuum observations with fully operational telescope

Contributors: Minh Huynh, Ray Norris, James Allison, Martin Bell, Dane Kleiner, Tao An

- Ingest rates - 500 MB/s currently, need to up to 200 TB/day for full ASKAP ops (so up ~6x). Will Pawsey upgrade be sufficient? Can ASRCs initial infrastructure development help?
- There are yet many issues with the noise in images being the biggest of all. It's currently higher than required for EMU, and 4 times above thermal. This is the biggest issue. We need to bring this down. Flagging, calibration, and the beam model need to be improved.
- Quality control. ASKAP has automatic continuum image check of positions, flux densities, source counts, etc. (Collier et al.) useful. What we need are the validation metrics to automate the QC.

EMU

Original EMU spec - one single observation. ASKAPsoft will produce a single continuum image from a single SBID. If EMU needs to combine data from different SBIDs, they currently need to do this offline by themselves as part of ASKAP Early Science work. For example, it is likely going to need in combining A and B interleaves from different days. In full ops having the ability image multiple SBIDs would be required. If this can't be done in ASKAPsoft, it'll have to be postprocessing.

ACES

The team is focused on science. (Postdocs need to publish). There's a need for more people dedicated to commissioning, who work on specific tasks (e.g. beamforming, bandpass calibration, etc.), and not several people who work on many bits. Also issues with people leading aspects and then leaving (e.g. J. Marvil). Where and how do we keep the expertise? Can ASRC be a solution?

VAST

The aspirations are to have 1 sec time scale continuum images for pular and AGN (scintillation science). It'll probably take 3 years. In year 1, if field is revisited after 12 hours that is okay. So daily images okay to begin. For images every 1 sec (this was in original spec), 300 x 1 MHz, 1 sec integrations, 36 antennas.
The spectral line team is currently doing 16,000 channels (or images) x 7 sec integrations over 12 hours. So this seems a tractable problem: shrinking in freq domain (300 channels vs 16000) but going to 1 sec integration (images) over 12 hours.

Lessons learned from the precursors that need to be taken in account in developing Australian SRC

Contributors: Attila Popping, Chenoa Tremblay, Tobias Westmeier, Minh Huynh, George Heald, James Allison, Ray Norris, Dane Kleiner, Bart Pindor, Katie Jameson, Chen Wu,

Martin Meyer, Youling Yue, Randall Wayth, Slava Kitaeff, Natasha Hurley-Walker, Tao An, Ray Norris

Australia hosts two SKA precursor telescopes: MWA and ASKAP. There were 3 rounds of discussions in this section. Below is the summary of those discussions.

Planning and Organisational lessons

- ASRC should not just design around the final product but also anticipate commissioning/early science phase. A lot of the current (ASKAP) design is for full ASKAP operating smoothly. In practice there is a phase of commissioning and early science with a lot of debugging, reprocessing and user access. Hence:
  - allow flexibility
  - build in commissioning requirements
  - have user quota and project policies
- Experts are the key to success. ASRC should absorb some of the experts that worked on the precursors to provide guidance in development.
- Documentation is essential. Proliferation of disconnected documenting tools and spaces is a problem. SRC should provide the document management system to be used across all the projects.
- ASRC should facilitate regular meetings between stakeholders early on in the development process (e.g. Pawsey-MWA-ASKAP working group which has started only relatively recently).
- Multiple software engineers need to be available to the science teams, and the integration needs to be quite close.
- The bottleneck will be a lack of people to work on projects, data, software, etc.
- Employ software engineers working together with astronomers to develop software.
- At the beginning, people dedicated to debugging the telescope, data and software will be needed (“commissioning scientists”); postdocs for doing science will come later. Making postdocs commissioning the telescope is damaging to their career because they can not produce publications while being busy with technical tasks.
- In planning the work expect delays. Allowing contingency is important.
- SKA1-Low development can benefit from testing the pipelines on precursor data due to the similarity of the telescopes (MWA).
- ASRC needs to be ahead of SKA1 to be ready to received the data, and can initially have the focus on precursors support.
- It is important that the project timeline, prioritisation of requirements, and staffing levels match the actual needs.
- In order to attract funding, ASRC needs to define projects that benefit both precursors and SKA.
- ASRC should be involved in commissioning of SKA1 in order to develop the knowledge of the telescope and be on a position to assist the science teams in planning the observations and post-processing.
- ASRC could be the central point in facilitating the exchange of expertise between stakeholders (SKA, universities, users, etc.), e.g. through organising busy weeks.

- High-level agreement and documentation on data access, responsibilities, interfaces, etc., would be beneficial, e.g. coordinated by ASRC coordination group, to secure funding for activities.

Technical lessons

- Data management tools are important. They should be user-friendly, and stress tested.
- System administrators should be part of the development teams to decrease the lag when a problem needs to be resolved at the system level of the system needs to be profiled or optimised.
- Having a test-suite for software upgrades while testing the proposed system during the procurement is important.
- Agile development should be widely adopted due to the fact that user requirements are not well known or can change.
- ASRC should optimise the system architecture for fast and convenient way of staging data from archive to HPC. Read/write from tape is slow and can be a great loss of efficiency.
- ASRC should have the procurement that is based on user requirements, otherwise there's a risk to get a system that is suboptimal. E.g., the current system at Pawsey has the storage and compute separate, which is not ideal for the data intensive applications. The current example at Pawsey - global filesystem does not work, instead it should have a dedicated astro filesystem with isolation to other groups. The systems need to be evaluated with a realistic test-dataset and the software before purchasing.
- Usage of the data storages (disk and tape) should have use policies from Day 1.

Development lessons

- New software will not work right from the start - software commissioning is a long process.
- The users and developers should have access to a container or virtual machine to customise and develop code before actual deployments. If this code works within the container, it should work on the HPC.
- Software should be scalable and work on different environments without changing the code. The user should focus on the methods and algorithms, rather than implementing the most effective use of nodes and cores etc.
- If there are pipeline changes, they are not "real-world" tested. There should be test-cases, to be provided by the science team, to validate system changes.
- ARC should provide workload optimisation services: a profiling and characterisation service. If a user wants to run something it will make prediction of the cost/time etc, and how to do this best.
- There should be a more flexible and "modular" pipeline to be able to make tweaks and changes easily.

Discovery the unexpected

Contributors: Ray Norris, Youling Yue, Tobias Westmeier, Bi-Qing For, Ian Bird, Peter Quinn, Ivy Wong, Ron Ekers, Antonio , Katie Jameson, James Allison, Ryan Shannon

Discovery of the unexpected is something you have to plan for.

What are conditions to make the discovery possible:

- Looking at data in different ways than originally planned;
- Don't throw away data, any data, ever. Tapes become cheaper every year;
- Provide access to resources for public and students;
- Open access to data;
- Give people time/freedom to try out other ideas;
- Create environment supportive of unconventional ideas;
- Foster cross-disciplinary collaboration;
- People need to understand their instruments very well;
- Exploration of full parameter space;
- Efficient tools to explore the data;
- Allow projects with high risks and wild imaginations;
- Try to limit assumptions in pre-processing;
- Develop culture of curiosity and innovation;
- Target diversity of community and thought;
- Value your "failures". Null-results tell you something;
- Check the "spam folder";
- Good access to multi-wavelength data.

What needs to be done within SRCs:

- Build machine learning tools and robots for explore data, mine unexplored data products detects anomalies;
- Build good tools for identifying artefacts & RFI;
- Build good models and big simulations to understand data;
- Provide long-term curation of data, metadata, documentation.

SKA1

ASRC support for SKA1-Mid science in Australia

Contributors: Martin Meyer, Youling Yue, Tobias Westmeier, Tao An, Lister Staveley-Smith, Andreas Wicenec, Bi-Qing For, Markus Dolensky, Katie Jameson, Dane Kleiner, Attila Popping, Ryan Shannon, Bart Pindor

Australian groups are involved in SKA1-mid on:

- HI surveys,
- AGN,
- Pulsars,

- VLBI[2].

SKA1-mid SDP assumptions that determine boundaries of SRCs

The main assumption is that SDP will deliver cleaned, continuum-subtracted data cubes and representative PSFs, however this may not be achieved initially. It's unclear at this stage if source-finding is in- or out-side of SDP.

Data

- One SKA1-mid data cube is ~170TB.
- Data products we must keep[3]:
  - Representative uv-data to ensure pipeline parameters are always optimal;
  - Intermediate products to calibrate pipeline;
  - Stacked uv grids;
  - PSF cube;
  - Multiple resolutions/weighting.
- Raw, intermediate data products:
  - Should always be able to export raw data from SDP;
  - Should be able to export intermediate data products from SDP;
  - Should be input/export pipelines from SDP/SRC.
- Should have ability to combine SKA1-MID catalogue with SKA1-LOW for foreground subtraction.
- Ability to calibrate SKA1-low beams from SKA1-mid (assuming sufficient frequency proximity).

Pulsars and FRB

- Should be able to access phase calibration solutions derived from imaging for fast transients.
- Pulsar timing: should be able to combine calibrated data products from SKA-mid and SKA-low.
- Pulsar timing algorithms have become data intensive so will need serious compute resources at ASRC.

Multimessenger/Multiwavelength

- ASRC should enable joint analysis across SKA1 image and catalogue products (SKA1-low & mid).

Early Science/Commission

- ASRC should assist with the development flexible pipelines and analyses data early on to enable tuning the final pipelines, and interact with modular pipelines.

---

[2] Should consult this community to determine their requirements.
[3] Also see ASKAP requirements document for Wallaby and DINGO.

Required support from ASRC

- Because the science requirements continue being refined, ASRC should continue having the input on the nature of the regional centre in coming years.
- ASRC should have an operational model. Potential model include eg having dedicated SRC resources for post-processing for any approved KSP.
- ASRC should have provision for software support for SKA1-mid projects to execute required post-processing in SRC.
- ASRC should support the work with SKA1-mid on equal basis to SKA1-low as many of Australian science groups are interested in the facility, and most of the HPSOs.

Notes on post-processing of continuum surveys with SKA1-Low

Contributors:

Fully automated pipeline processing in SDP is unrealistic. Community will need access to computational resources to post-process and improve the pipelines.

There are two types of post-processing:

1. On images, that include:
    a. Source extraction
    b. Stacking
    c. RM Synthesis
    d. Variable/transient detection
    e. Moment analysis
    f. Spectrum extraction
2. On catalogues
    a. Cross-identification
    b. Variable/transient detection
    c. SEDs / spatial fits
    d. Number counts

General needs

- Sufficient user disk space to run experiments and execute post-processing;
- Good software/tools/packages management that users don't have to do it themselves;
- Scalable and flexible platform;
- The platform with good I/O, because the data is big;
- Effective remote visualisation (should not require transporting the data to visually inspect it);
- Programmatic access to the data in archive from pipelines;
- Well curated data;
- VO compliant facilities;
- Good user support at all levels;
- Well documented standard analysis tools;

Post-processing needs at ASRC of spectral-line surveys with SKA1-Low

Contributors: James Allison, Chenoa Tremblay, Dane Kleiner, Chen Wu, Slava Kitaeff

SKA1-Low will have significant capabilities for the galactic and extragalactic spectral line observations.

Science goals

- HI and molecular absorption lines (high redshift quasars and radio galaxies - outflows & gas accretion).
- Bio-molecular lines (cradle of life science).
- Recombination lines (physical conditions of the interstellar medium).

Data requirements

- Ideally SKA (but probably not realistic initially) would provide flagged, fully calibrated and continuum-subtracted visibilities.
- Minimal requirements are (a) perfectly continuum-subtracted cubes at full spectral resolution (b) continuum cubes / models spanning full bandwidth.
- Spatial resolution / PSF of cubes depends on science question. Should provide a discrete range of cubes at different spatial resolutions optimally matched to “all” possible science questions.

Post processing

- Source finding and line fitting in cubes with ability to use priors.
- Generation of moment maps from cubes.
- Ability to assess quality and feed this back to SDP.
- Capability to visualise cubes and post-processed science products.
- Perform operations on cubes - slice in spatial and spectral domain, measure key parameters such as velocities, widths, equivalent widths (combine with continuum model), line fitting, deblending.
- Catalogue requirements & operations:
    - Cross match against other major surveys at other wavelengths;
    - Data mining operations online catalogues;
    - Provide VOSpace;
    - Support use of “standard” tools used by wider non-radio-expert community, e.g. ALADIN, SKYVIEW.

Post-processing for pulsars search and timing at ASRC

Contributors: Ryan Shannon, Youling Yue, Phil Edwards, Chen Wu, Markus Dolensky

Pulsar discovery workflow: Searching -> Confirmation -> Timing

Pulsar searching

- We can’t save the raw data due to the large size, but we will save the candidates.

- SDP provides a list of ranked candidates. These candidates need to be confirmed through the observations that need to be scheduled. Can the observations be automatically triggered through SRCs?
- The candidates need to be permanently archived.
- There's a need in an effective visualisation tool to work with the candidate database.
- SRC should provide a capability to cross match with different catalogues.

Pulsar software

- Advanced Bayesian algorithms for timing analysis (moving from converting profiles into arrival times towards analysis directly on the pulse profiles).
- GPUs accelerated codes.

Timing

- The data is relatively small.
- We need to be able to combine SKA-low and SKA-mid data. SKA-low data will provide the dispersion measure corrections for SKA-mid searches.
- If glitches and binary pulsars are detected, the corrections need to be fed back into SDP. There's a need for a programmatic link SRC->SDP SKAO.
- SRC should allow sophisticated analysis such as GW searches or binary characterisation.

SKA related theory and simulations work that should be supported by ASRC

Contributors: Chris Power, Minh Hyunh, Tao An, Randall Wayth, Martin Bell, Ryan Shannon
We identified broad requirements of theory and simulations;

- Physically-motivated galaxy formation and evolution predictions, including e.g. radio continuum emission associated with AGN and star formation in galaxies, that can be compared directly with observational datasets.
- Physically-motivated, testable predictions for the properties of transients, incorporating not only stellar population modeling to derive e.g. rates but also the microphysics associated with generation of the signals and features in the emission spectra.
- Simulated skies incorporating instrument and telescope effects, and either semi-empirical statistical realisations or physical models of astrophysical sources to facilitate design and planning of surveys, to maximise the power of the surveys to discriminate between competing theoretical models.
- Modelling of the Galaxy, both as a foreground and as a host of interesting source populations.

The techniques for creating synthetic galaxy populations in cosmological volumes with physically-motivated observables are both mature and sophisticated, and there is good reason to expect that the requirements of the cosmology and galaxy evolution communities can be satisfied with extensions of current approaches. There is also good reason to believe that these simulated galaxy populations can be combined with simulations that model instrumental effects in a reasonably straightforward fashion. Some thought should be given to having a uniform interface so that the source population input can be readily convolved with instrumental effects and analysed using the software typically used by the community.

The requirements of the transients community are more demanding. For example, the event rates and distributions of Galactic transients will have a direct impact on survey cadences, while the microphysics that needs to be modelled to understand how the signal can be inverted to inform our understanding of e.g. event energetics and geometry is extremely challenging to model. Stellar population models are reasonably mature, but it is not yet obvious what the variance between models is, and how e.g. extensions to binary population synthesis modeling will affect results; this will be crucial for predicting event rates. Statistical estimation of spatial distributions may very well be sufficient (e.g. correlating with a given type of stellar population), but modelling of the microphysics brings with it significant uncertainties, and it is not clear whether there is the expertise within the current Australian community.

Synergies with existing ARC Centres of Excellence (ASTRO 3D, OzGrav) should be exploited. Longer term, it is essential that theory, simulations, and modelling capability is incorporated into an ASRC to provide the technical (optimal survey design and strategy) and science support to maximise the scientific return on investment. A core team that has expertise in statistical modelling, instrument simulation, science simulation, and delivering mock data should be integral to future ASRC planning. A Tier 2 facility could support the kinds of simulations and data manipulation and serving to support the regional SKA community.

Recommendation: Set up a technical working group tasked with producing full end-to-end simulations of source populations (physical or semi-empirical) and instrumental/telescope effects to facilitate testing and analysis of the pipelines to be used in future.

Recommendation: Set up a technical working group focussed on the (micro)physics of transients. There may be overlap with existing efforts within ASTRO 3D (focused on stellar population modeling) and OzGrav (compact object mergers), as well as within the broader community, and this expertise shown be drawn in.

Recommendation: Specialist science simulations for SKA science require a dedicated team, most appropriately embedded within the ASRC, that can value add to the survey science teams. Collaborations with simulations and modelling groups within the wider communities should be leveraged, but dedicated expertise is required to provide the kinds of science simulations required by diverse radio astronomy surveys.

3. Later included contributions

Not all science groups were able to participate in ASRC Workshop. This section includes the contributions sent after the workshop.

EoR/CD Science

Author: Cathryn Trott (ICRAR-Curtin)[4]

This short document outlines the potential interface between the Science Data Processor and a SKA Regional Centre (SRC), and EoR/CD data being handed to Science Teams (outside of the SKA SDP structure, and inside and outside SRCs).

EoR/CD experiments:

There are three primary EoR/CD experiments:

1. Statistical detection and estimation - these are primarily a set of spatial power spectrum experiments undertaken across the full SKA1-Low bandwidth (50-350 MHz), encompassing the Cosmic Dawn and EoR. The higher redshifts are inaccessible to current experiments, and this forms new ground for SKA;
2. Direct imaging (21cm tomographic cubes) - mapping of the brightness temperature distribution of 21cm gas across redshift. This experiment requires deep pointings on well-calibrated and cold fields;
3. 21cm Forest - mapping of the neutral hydrogen structure along sight lines to distant continuum sources (high-z QSOs).

All three experiments share the common requirements of:

1. Long integration times
2. Accurate and precise calibration
3. Deep foreground source subtraction
4. Smooth spectral response of the telescope
5. Access to raw data with associated metadata describing the full instrument state and post-analysis steps.

This latter requirement demands an early exchange of EoR data from SDP to the science team, nominally through a SKA Regional Centre. Within this SRC, data will be further processed and treated by the EoR science team and staged to other SRCs internationally for access by other team members. Given the proximity of Perth to the SKA-Low site, the natural host location for EoR/CD data and science is the ASRC.

SDP EoR/CD Interface

The EoR/CD experiments require:

- Multi-stage calibration
    - Bandpass
    - direction-independent Jones matrices

---

[4] With contributions from Leon Koopmans (Kapteyn Institute), Sarod Yatawatta (ASTRON), Andreas Wicenec (ICRAR-UWA)

    - direction-dependent Jones matrices
    - incorporation of differing station beams
    - RFI excision, channel flagging, weighting
    - dynamic beam changes (e.g., dead dipoles)
    - spectral and temporal averaging
    - precise source peeling.
- Assessment, logging and uniformity of all steps applied to the data.

The SDP - EoR/CD interface lies early in this process, where basic calibration, RFI excision and temporal and spectral averaging have occurred.

ASRC Requirements

For EoR/CD, the local SRC plays a role between the raw-level data products and metadata provided by SDP, and the wider international EoR/CD community. In this guise, the SRC plays multiple key roles:

1. Housing/hosting of EoR/CD SWG members to provide the initial processing of data
2. Hosting of EoR/CD software for downstream calibration and processing
3. Storage of averaged visibility data and metadata
4. Staging of data products to the international community, via delivery to external SRCs
5. Interfacing with other science teams for sharing of science products