AUSSRC-DSP-2020-0002

AusSRC Radio Astronomy Data User Community Survey Report	
DESIGN STUDY PROGRAM	
Date:	**16.07.2020**
Work package:	**WP5**
Dissemination level:	**Public**
Published at:	**https://aussrc.org/documents/**

Abstract

Australian SKA Regional Centre (AusSRC) Design Study Program undertook a survey of the radio astronomy data user community in Australia for the purposes of obtaining a feedback that would further inform the development of AusSRC. The results based on seventy-four responses, conclusions and future actions obtained from the survey are detailed in this report.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	**Date**
From	S. Kitaeff	
Author(s)	L. Quartermaine, S. Kitaeff	
Reviewed by	P.Quinn. S.Pearce	
Approved by	P.Quinn	16.07.2020

III. DOCUMENT LOG

Issue	**Date**	**Comment**	**Author**
1	22.11.2019	Initial draft	Louisa Quartermaine
2	15.05.2020	Developed draft for discussion	Louisa Quartermaine Slava Kitaeff
3	1.06.2020	Final draft	Louisa Quartermaine Slava Kitaeff
4	16.07.2020	Release	Louisa Quartermaine Slava Kitaeff

IV. GLOSSARY

PI	Principal Investigator
ALMA	Atacama Large Millimeter/submillimeter Array
API	Application Programming Interface
ARC	ALMA Regional Centre
ASKAP	Australian SKA Pathfinder
ASVO	All-Sky Virtual Observatory
ATNF	Australia Telescope National Facility
CASDA	CSIRO ASKAP Data Archive
CSP	Central Signal Processor
ESA	European Space Agency
ESDC	European SKA Data Centre
ESRC	European SKA Regional Centre
EVN	European VLBI Network
FFT	Fast Fourier Transform
FOV	Field of View
GSM	Global Sky Model
GUI	Graphical User Interface
HPC	High Power Computing
HPSO	High Priority Science Objective
HST	Hubble Space Telescope
HTC	High Tech Computing
IVOA	International Virtual Observatory Alliance
JIVE	Joint Institute for VLBI ERIC
LBA	Long Baseline Array
LOFAR	LOw Frequency ARray
LSM	Local Sky Model
MWA	Murchison Widefield Array
NASA	National Aeronautics and Space Administration
QA	Quality Assessment
SDP	Science Data Processor
SED	Spectral Energy Distribution
SG	Science Gateway
SIAP	Simple Image Access Protocol
SKA	Square Kilometre Array
SRC	SKA Regional Centre
SSAP	Simple Spectral Access Protocol
TAP	Table Access Protocol
VLBA	Very Long Baseline Array
VLBI	Very Long Baseline Interferometry
VO	Virtual Observatory

VII. EXECUTIVE SUMMARY

AusSRC Design Study Program surveyed the radio astronomy data user community in Australia to obtain feedback that would further inform the development of AusSRC. The results based on seventy-four responses, conclusions and future actions obtained from the survey are detailed in this report. These results provide further detailing for the community requirements for the development of the Australian SKA Regional Centre.

Table of Context

1. Introduction

In November 2017, the first Australian SKA Regional Centre community workshop was held in Perth. The output of the workshop was summaries in two papers published on AusSRC website [RD1, RD2]. In 2019, in preparation for the second community workshop, this time held at Mt Stromlo (Canberra, ACT) in November 2019, the survey detailed in this report was proposed and conducted with preliminary results presented at the workshop for further discussion and consideration.

2. Goals of this document

This document aims to outline the results, conclusions and future actions obtained from the recent survey that will assist in informing the future Australian SKA Regional Centre (AusSRC) definition.

3. Survey Rationale

The aim of the survey was to build on the information provided by the previously collected data with a focus on current gaps in the services, support and infrastructure provided, and on how this information could be mapped onto the community expectations from the AusSRC. The results of the survey inform and update the definition and requirements of the future AusSRC.

This information will be of significant value in further updating the AusSRC white paper or developing a new paper.

4. Survey Structure

In designing the survey, we aimed at sampling all important aspects of doing science with radio astronomy data. The idea was to profile the user, tools, data, and to take a snapshot of the field now in order to repeat it two years to see what has changed.

Thus, the survey contains the following six sections:

1. ***User Profiling*** **-** this survey section sought information about a user of radio astronomy data, such as areas of scientific interest, position and role(s) and skill level working with radio data as well as various computing skill levels.
2. ***Projects Profiling*** **-** this survey section sought information on the projects with which users are involved, such as type of projects, nature of the collaboration, radio astronomy facility used. In this section we have also tried to identify some of the common observational project related challenges.

3. ***Data Profiling*** **-** this survey section sought information about the data and some specific challenges related to it.
4. ***Software and Tools Profiling*** **-** this survey section sought information on the type of software and tools used to process and interact with radio astronomy data, such as data reduction, statistical analysis, cross matching of catalogues, visualisation and image processing, as well as identifying the issues encountered when using those software and tools.
5. ***Data Processing Profiling*** - this survey section sought information on where users process their data, what type of data manipulation and interaction they perform, their satisfaction with data and compute facilities used, as well as the issues they encountered when processing data.
6. ***Expectations from AusSRC*** - this section requested users to rate the importance of what AusSRC should be providing in the future.

5. Method of Conducting the Survey

The survey was loaded into Qualtrics survey software platform kindly provided to us by the University of Western Australia. The survey contained 33 questions in total, and was conducted fully online. The link to the survey was sent out to the members of the Astronomical Society of Australia, twice, 6 weeks apart. Seventy-four survey responses were received, which represents approximately 60 to 70% of the member of Australian radio astronomy science community.

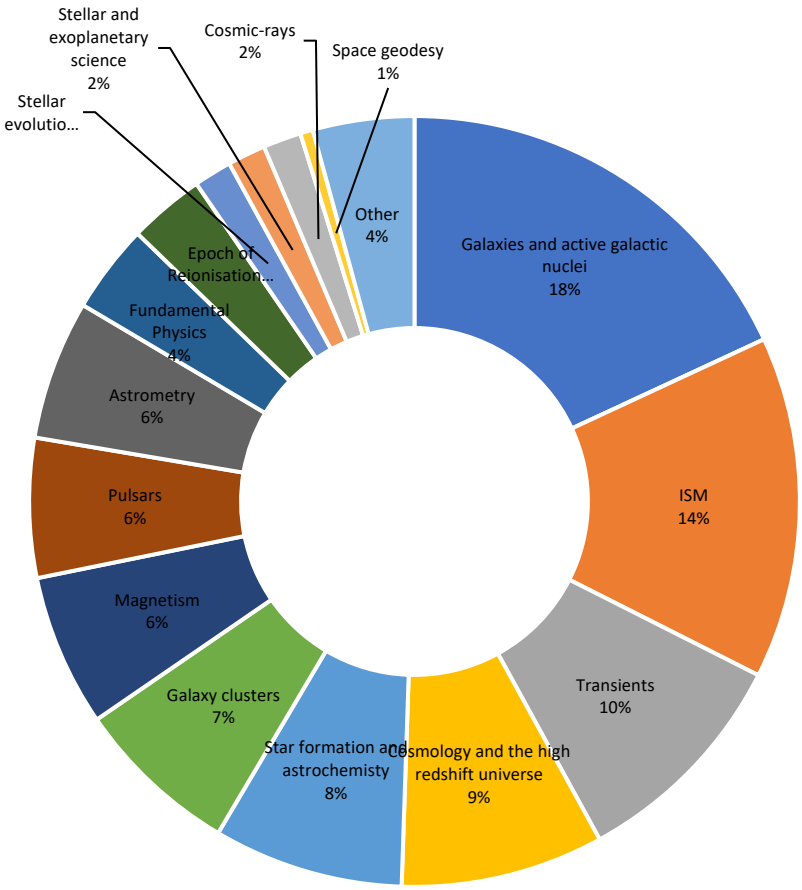
6. Survey Results

This section of the report details the results and key observations for each of the six-survey sections described in 4.

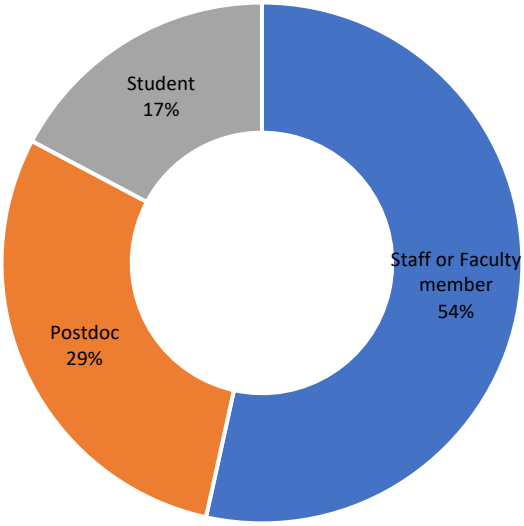
6.1. User Profiling

This survey section sought information about a user of radio astronomy data, such as areas of scientific interest, position and role(s) and skill level working with radio data as well as various computing skill levels.

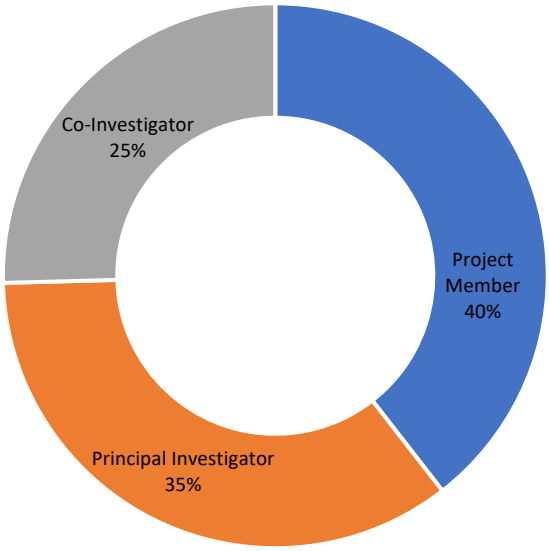
6.1.1. What are your areas of scientific interest



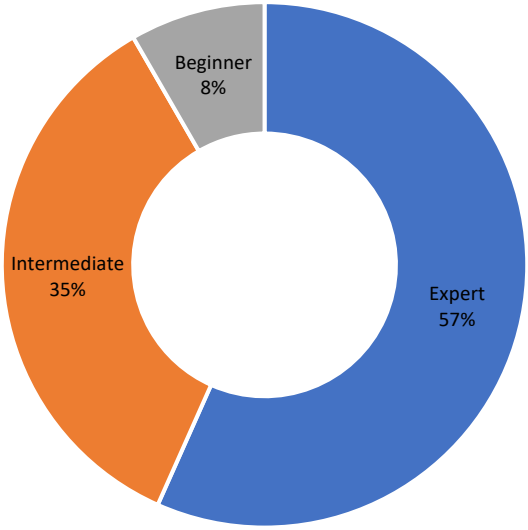
6.1.2. What position do you hold?



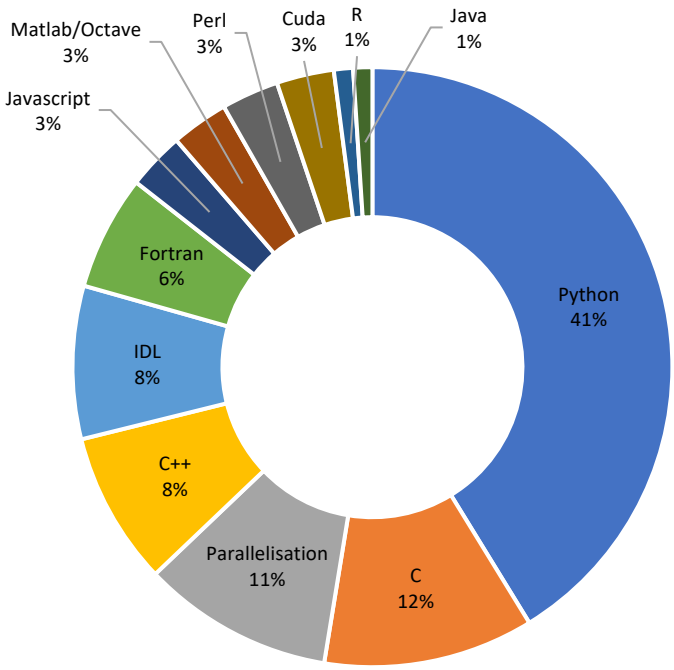
6.1.3. What project role do you hold?



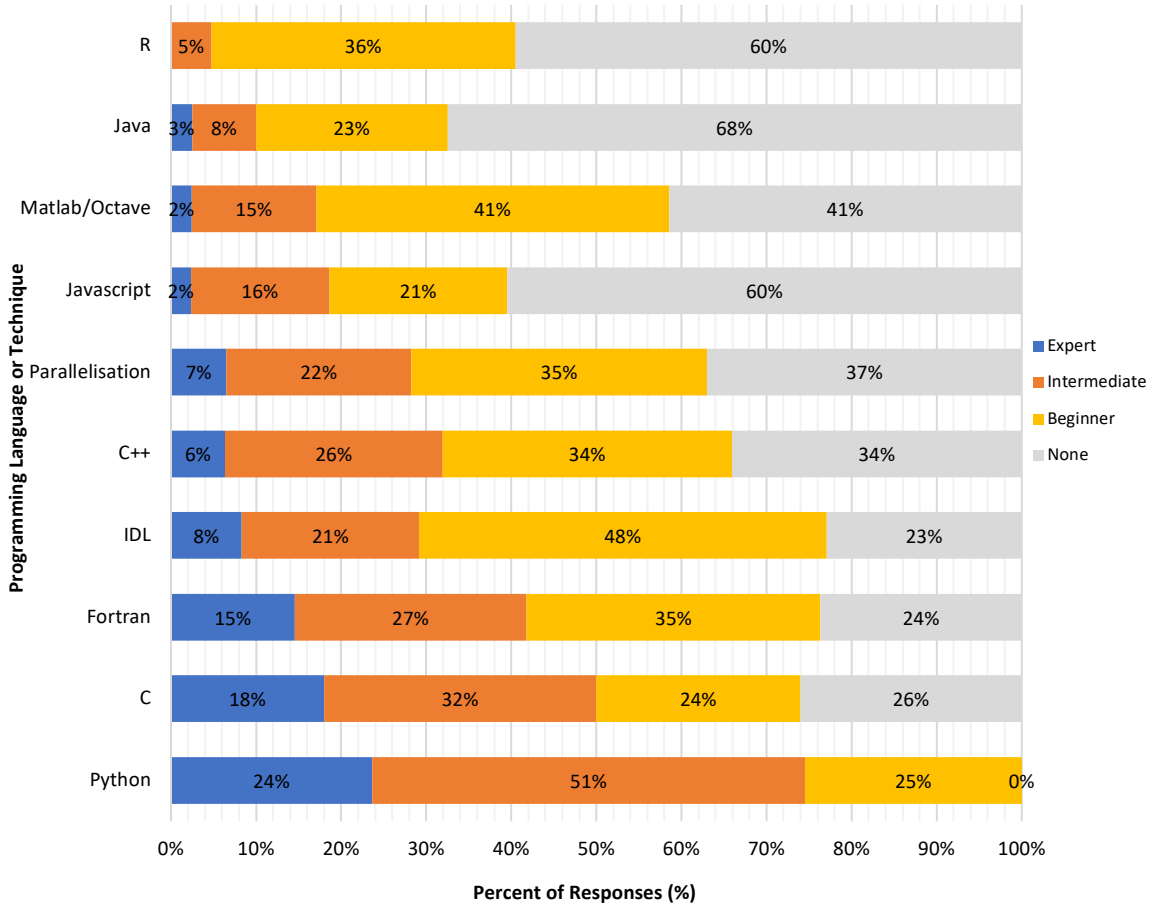
6.1.4. What is your skill level working with radio astronomy data?



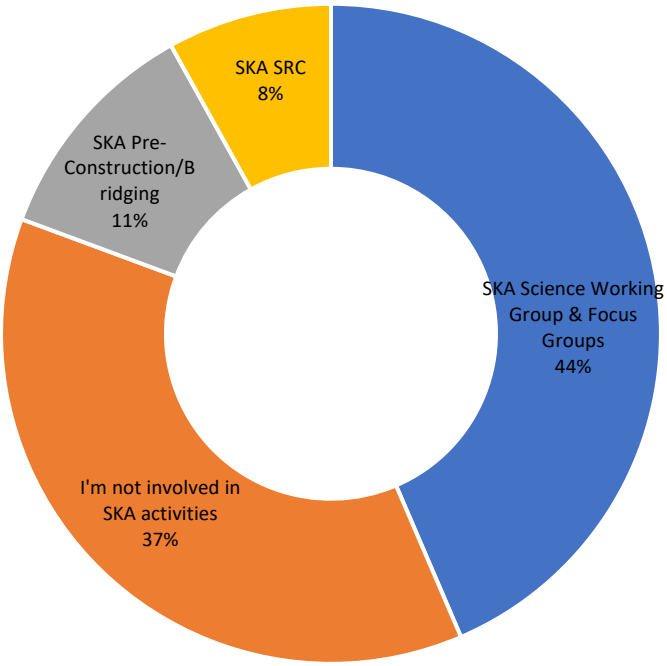
6.1.5. Which programming languages and techniques do you use often?



6.1.6. What is your skill level with these languages and techniques?

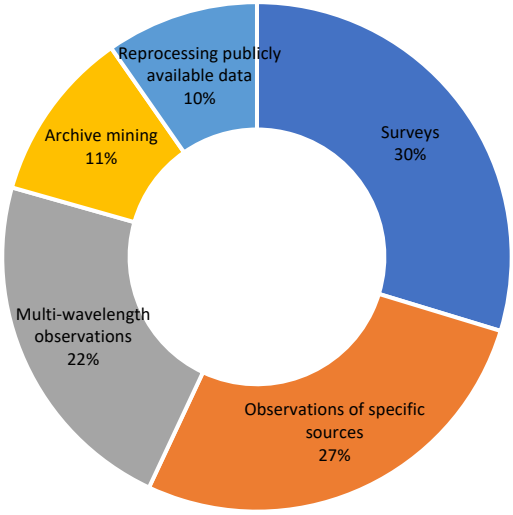


6.1.7. Do you participate in any SKA related activities?

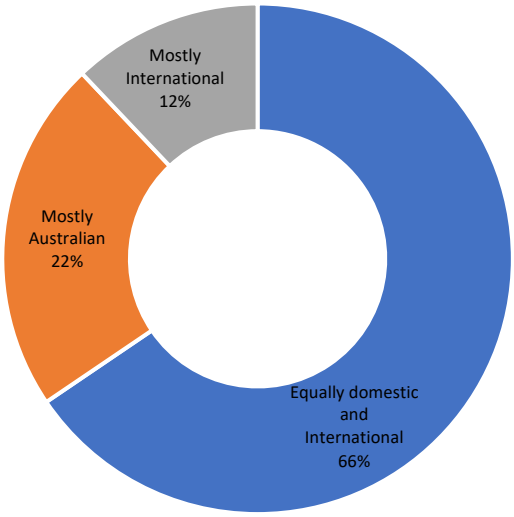


6.2. Projects Profiling

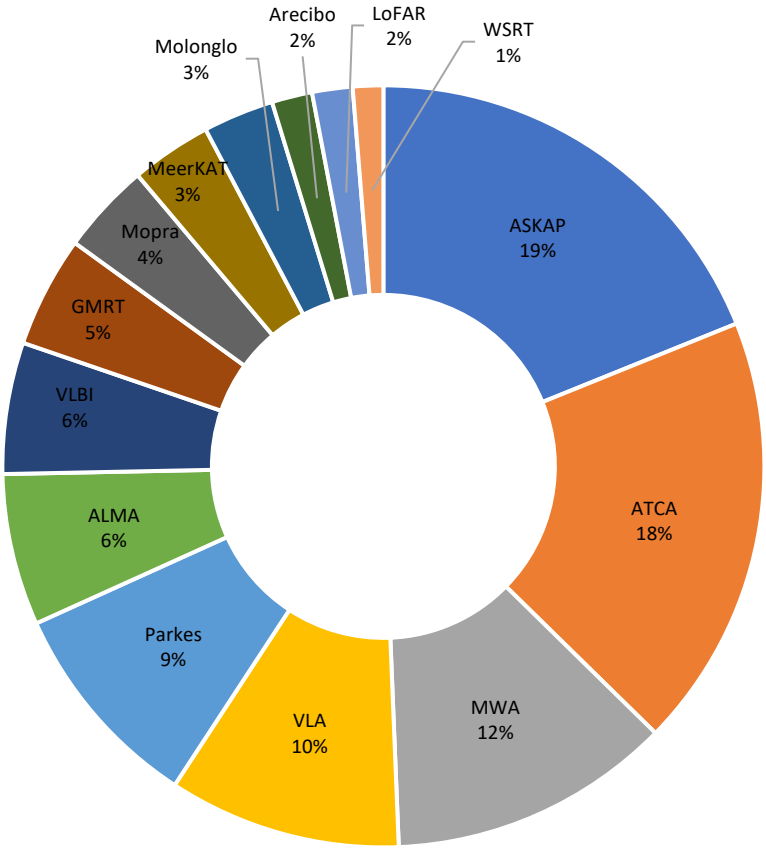
6.2.1. With what types of projects are you currently involved?



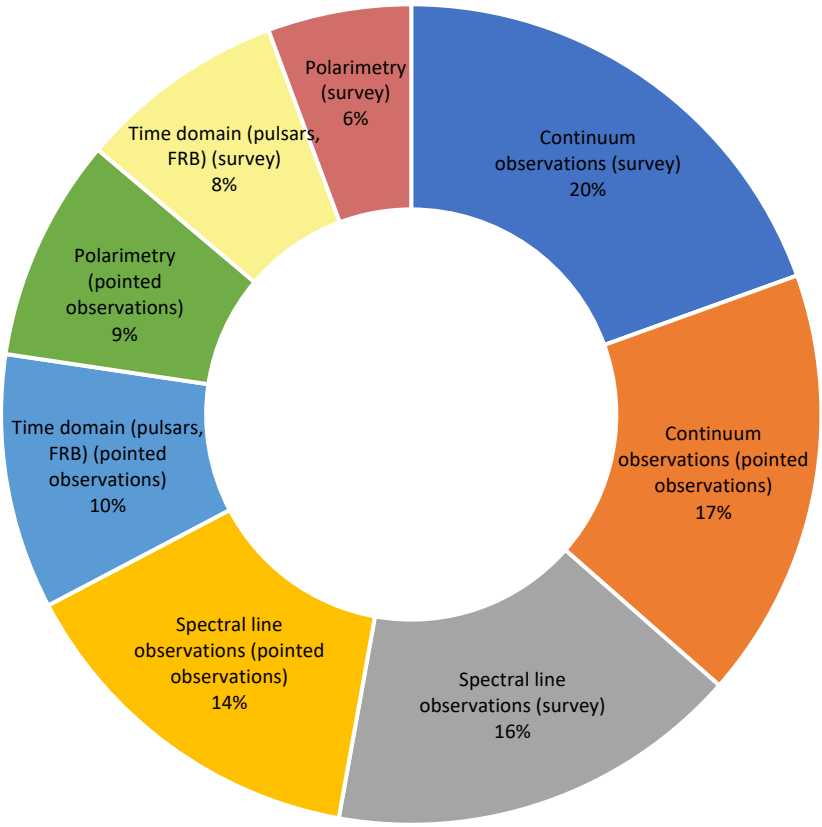
6.2.2. What is the nature of your current collaborations?



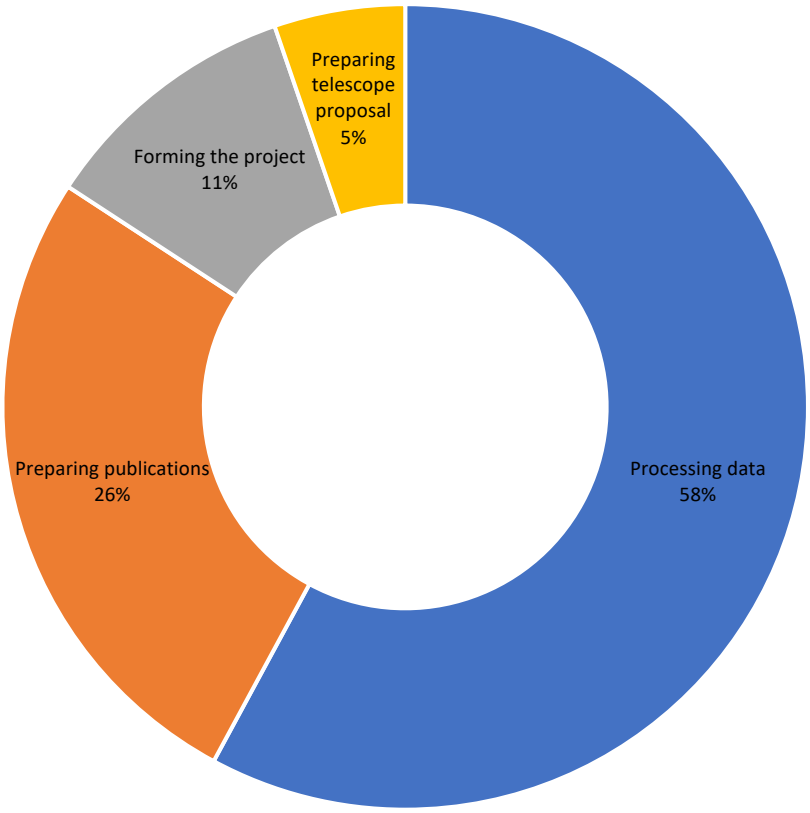
6.2.3. Which radio astronomy facilities do you currently use?



6.2.4. What is the observational nature of the data you are using?

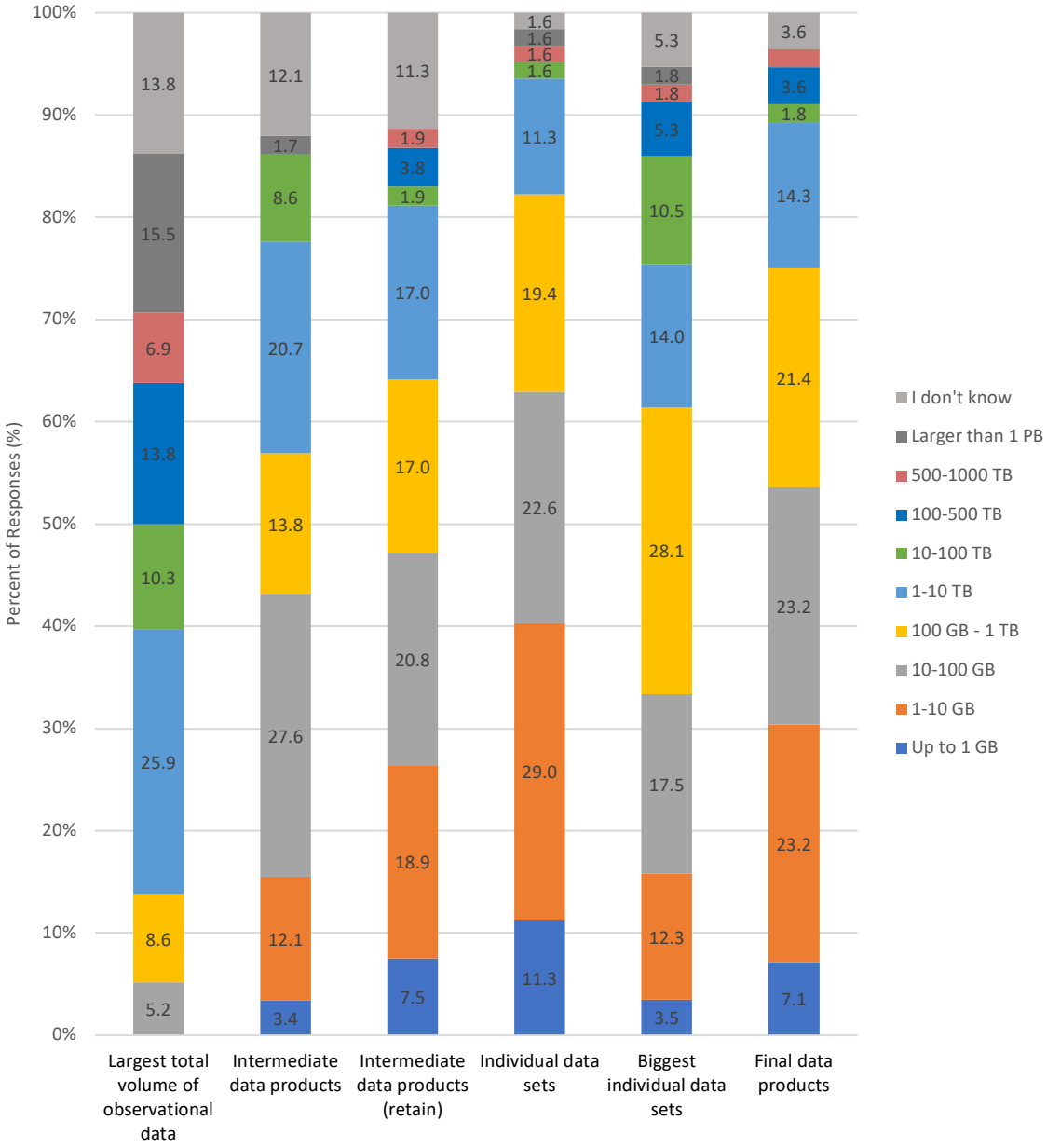


6.2.5. What is the most challenging stage in observational projects?

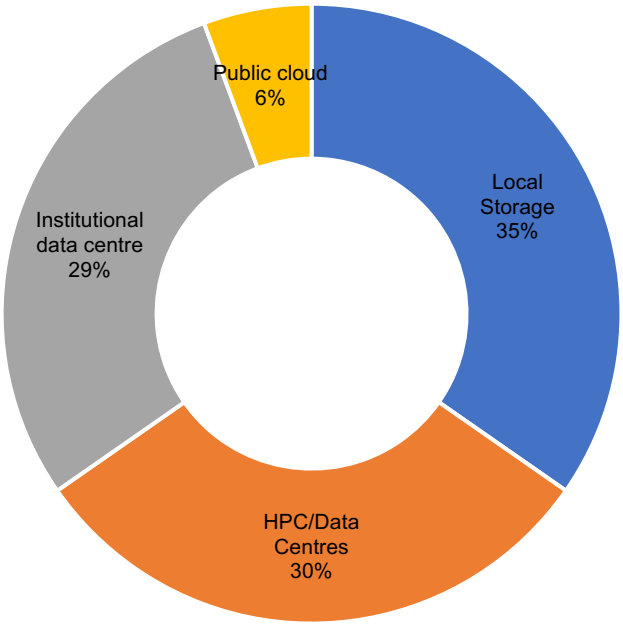


6.3. Data Profiling

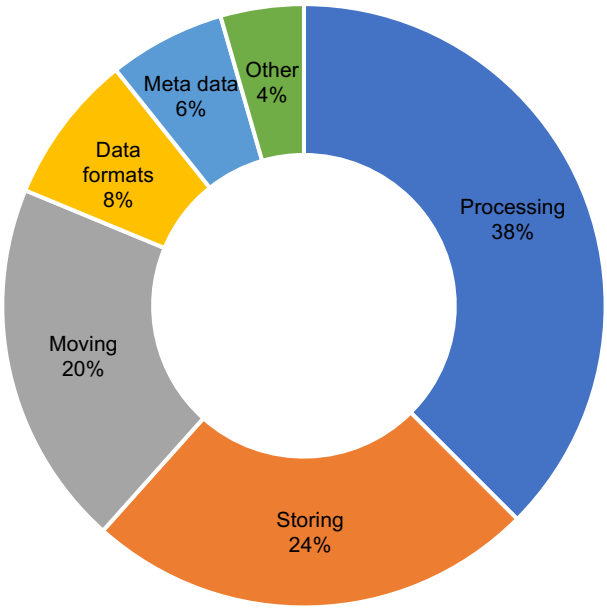
6.3.1. What is the size of the data with which you interact?



6.3.2. Where is your data stored?



6.3.3. What are the biggest issues you encounter when working with radio data?



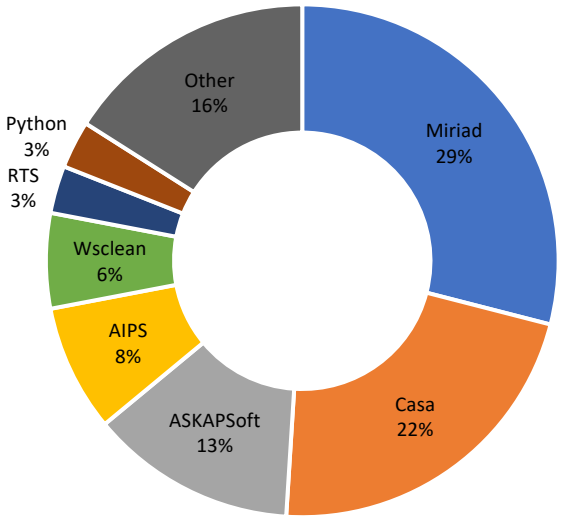
6.3.4. Biggest issues working with the data – elaborated

The respondents were asked to elaborate on what they considered as the biggest issues encountered when working with radio astronomy data. The individual responses can be summarised into the following most common issues:

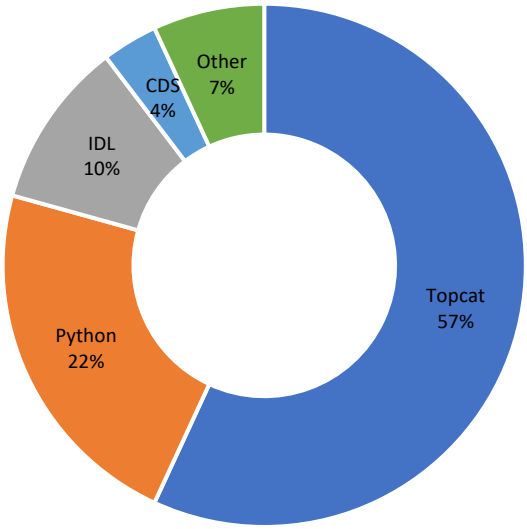
- The large volume of data requires the use of supercomputers which have become bottlenecks in the areas of processing, queue lengths, disk space, retrieving data and are considered a significant impediment to scientific research;
- The large amounts of data require more storage which is often not available, is limited and is not long lived;
- Moving and accessing large amount of data is problematic - time consuming and challenging;
- Manipulating, interacting and visualising large data sets can be problematic;
- The lack of meta data is a problem for querying archives;
- There is a lack of information available on the quality of data which leads to labor intensive and time-consuming manual inspection of data.

6.4. Software and Tools Profiling

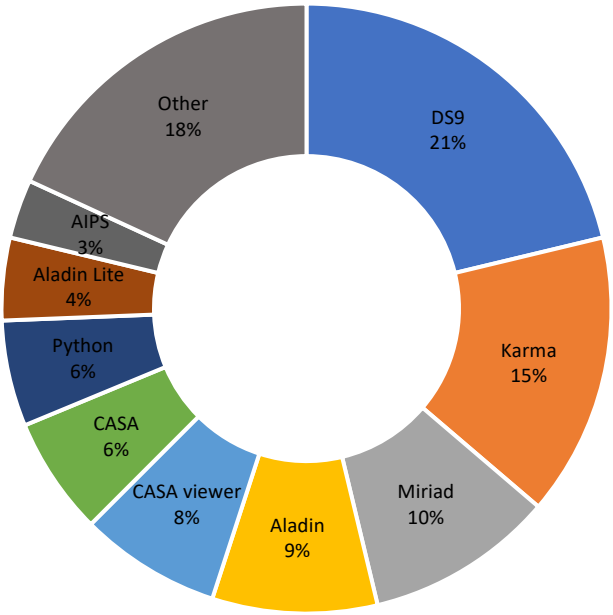
6.4.1. What data reduction software do you use?



6.4.2. What software do you use to work with catalogues for statistical data analysis and cross matching?



6.4.3. What software do you use for visualisation and image processing?

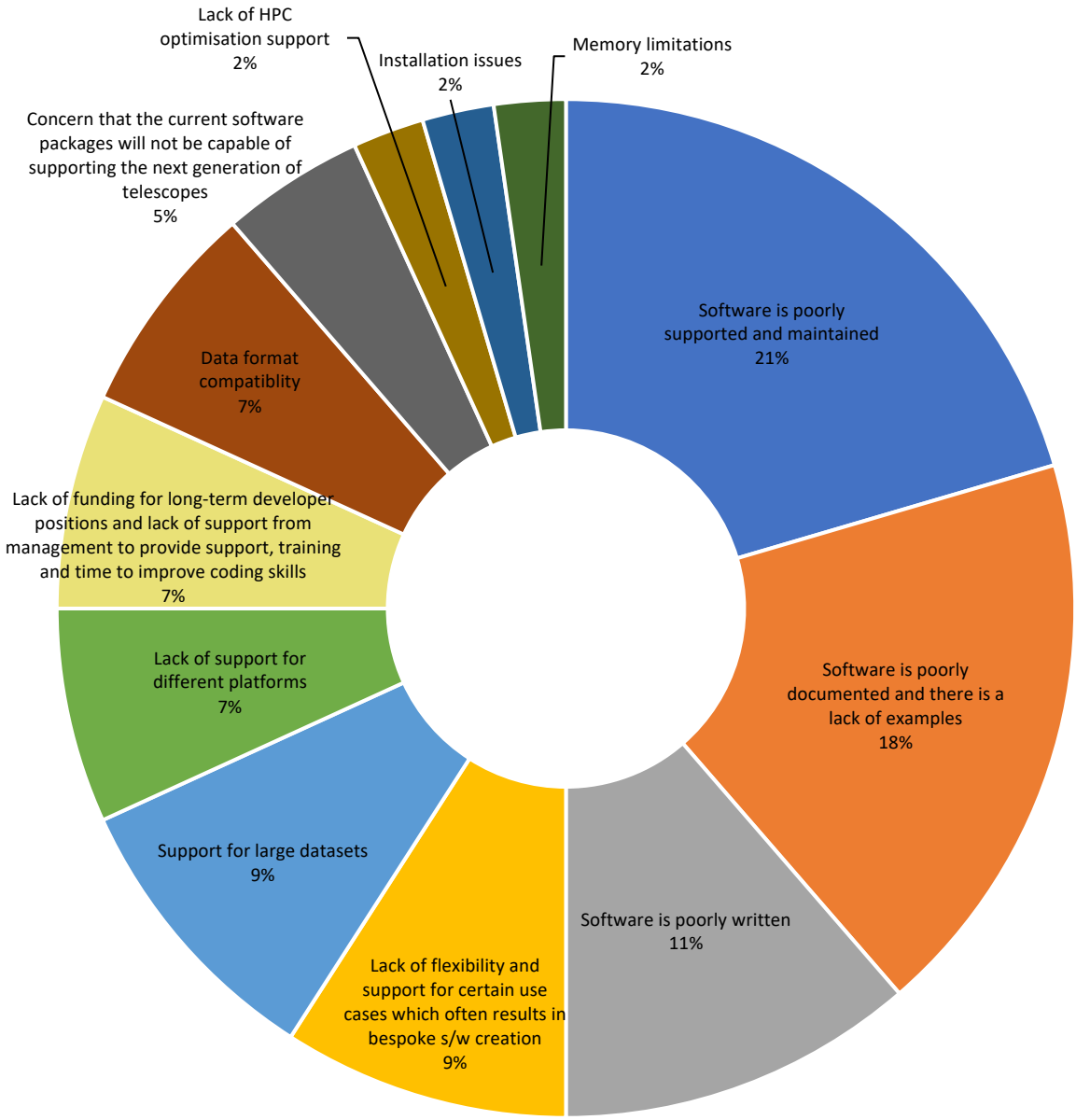


6.4.4. What are the biggest issues that you encounter using astronomy software and tools?

The respondents were asked to elaborate on what they considered as the biggest issues encountered when using astronomy software and tools. The individual responses can be summarised into the following most common issues:

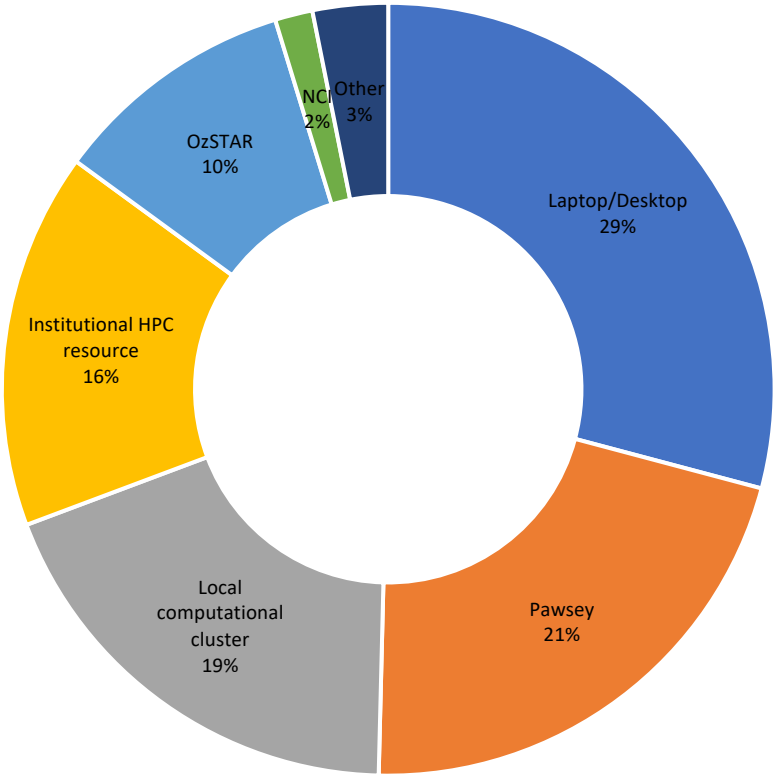
- Software is poorly written;
- Software is poorly documented and there is a lack of examples;
- Software is poorly supported and maintained – often issues are encountered when upgrading to newer versions;
- Concern that the current software packages will not be capable of supporting the next generation of telescopes;
- Lack of support for different platforms;
- Lack of flexibility and support for certain use cases often results in bespoke software creation;
- Lack of HPC optimization support;
- Lack of funding for long-term developer positions and lack of support from management to provide support, training and time to improve coding skills.

The following graph shows the distribution of the responses recorded in the list above.

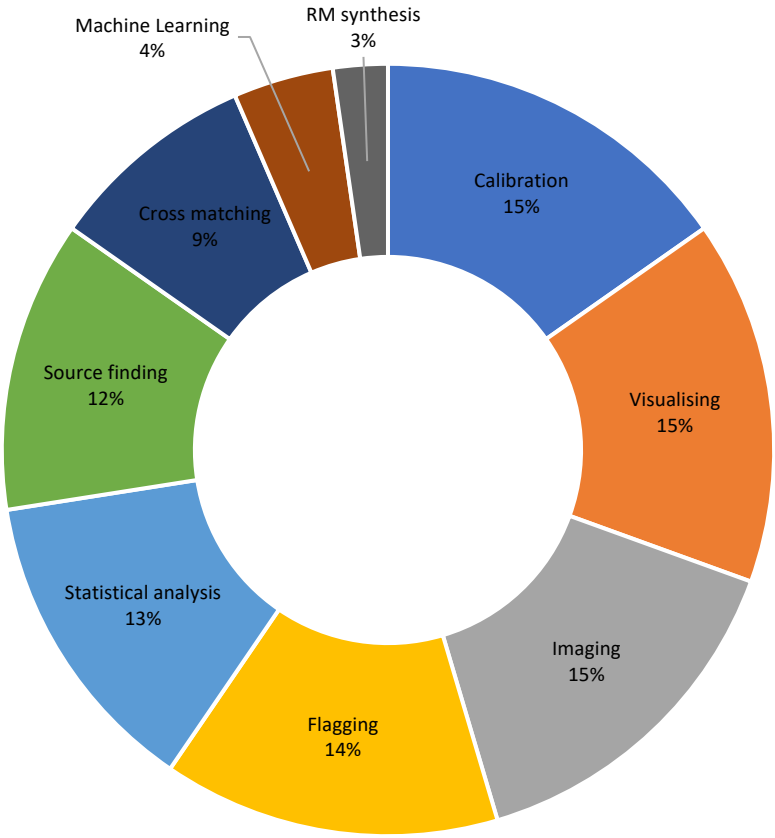

Lack of HPC optimisation support 2%
Installation issues 2%
Memory limitations 2%
Concern that the current software packages will not be capable of supporting the next generation of telescopes 5%
Software is poorly supported and maintained 21%
Data format compatiblity 7%
Lack of funding for long-term developer positions and lack of support from management to provide support, training and time to improve coding skills 7%
Lack of support for different platforms 7%
Software is poorly documented and there is a lack of examples 18%
Support for large datasets 9%
Software is poorly written 11%
Lack of flexibility and support for certain use cases which often results in bespoke s/w creation 9%

6.5. Data Processing Profiling

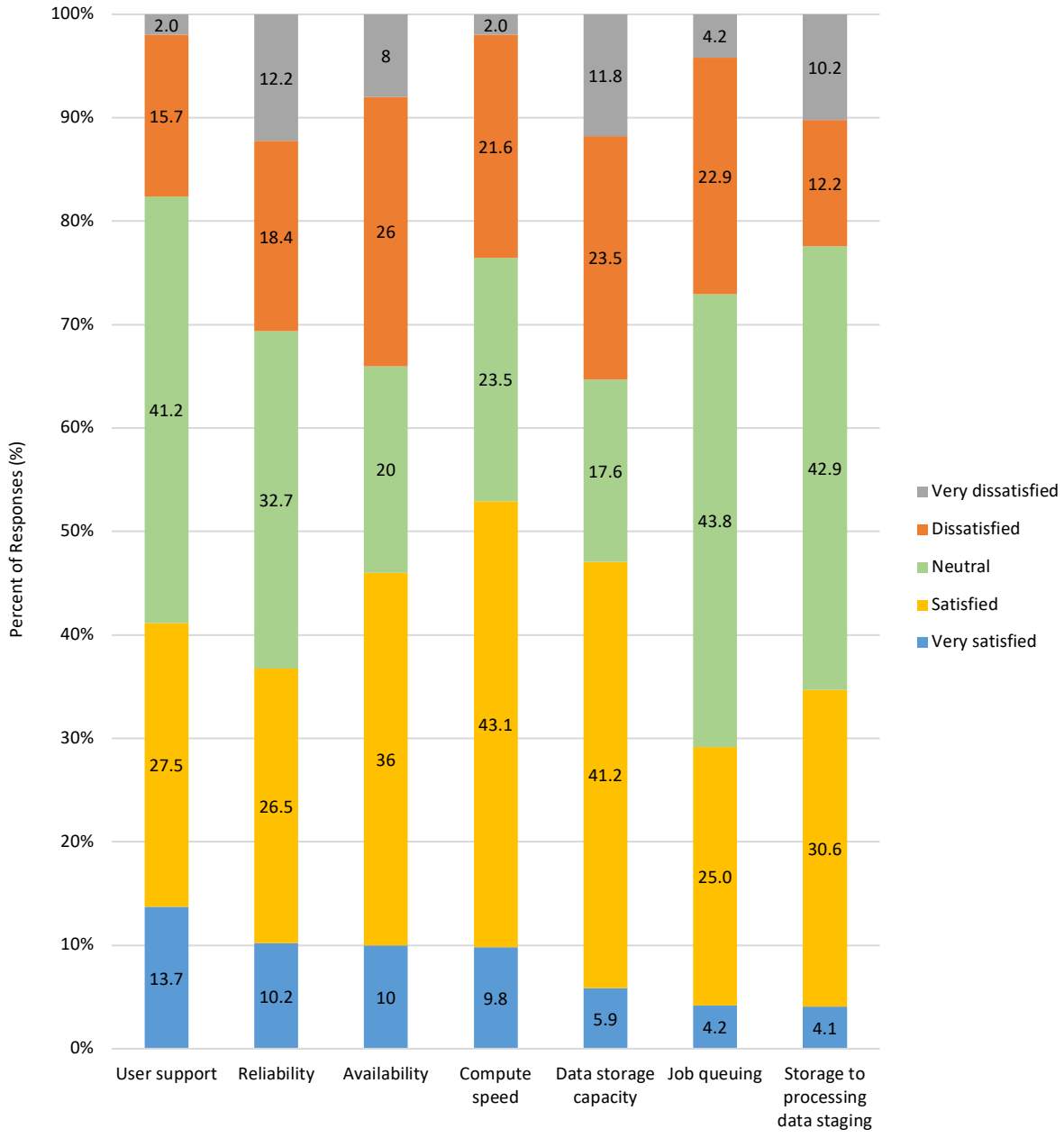
6.5.1. Where do you process your data?



6.5.2. What type of data manipulation/interaction do you perform?



6.5.3. What is your level of satisfaction with the data and compute facilities you use?

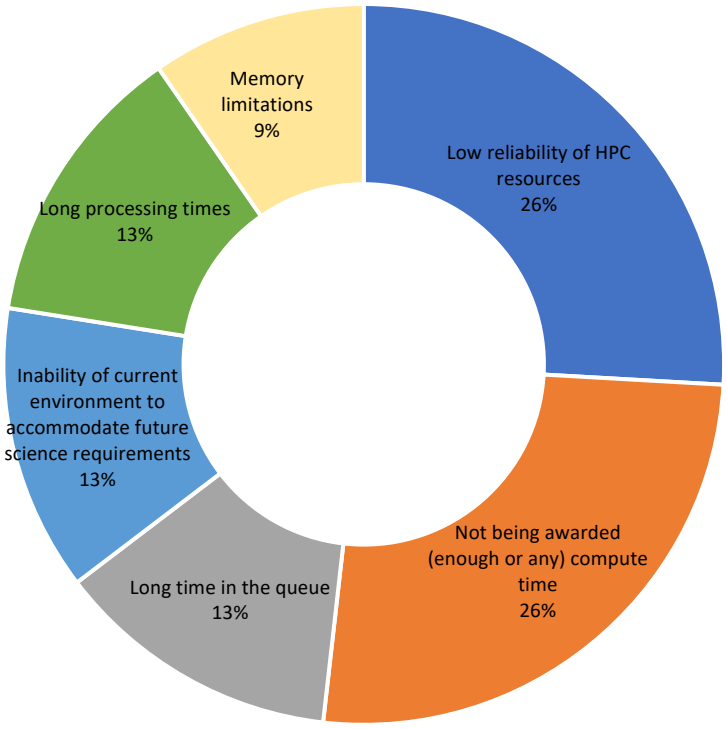


6.5.4. What are the biggest issues you encounter when processing the data?

When asked to provide further details on what respondents considered the biggest issues encountered when processing the data, 30 individual responses were received. These responses can be grouped into the following categories:

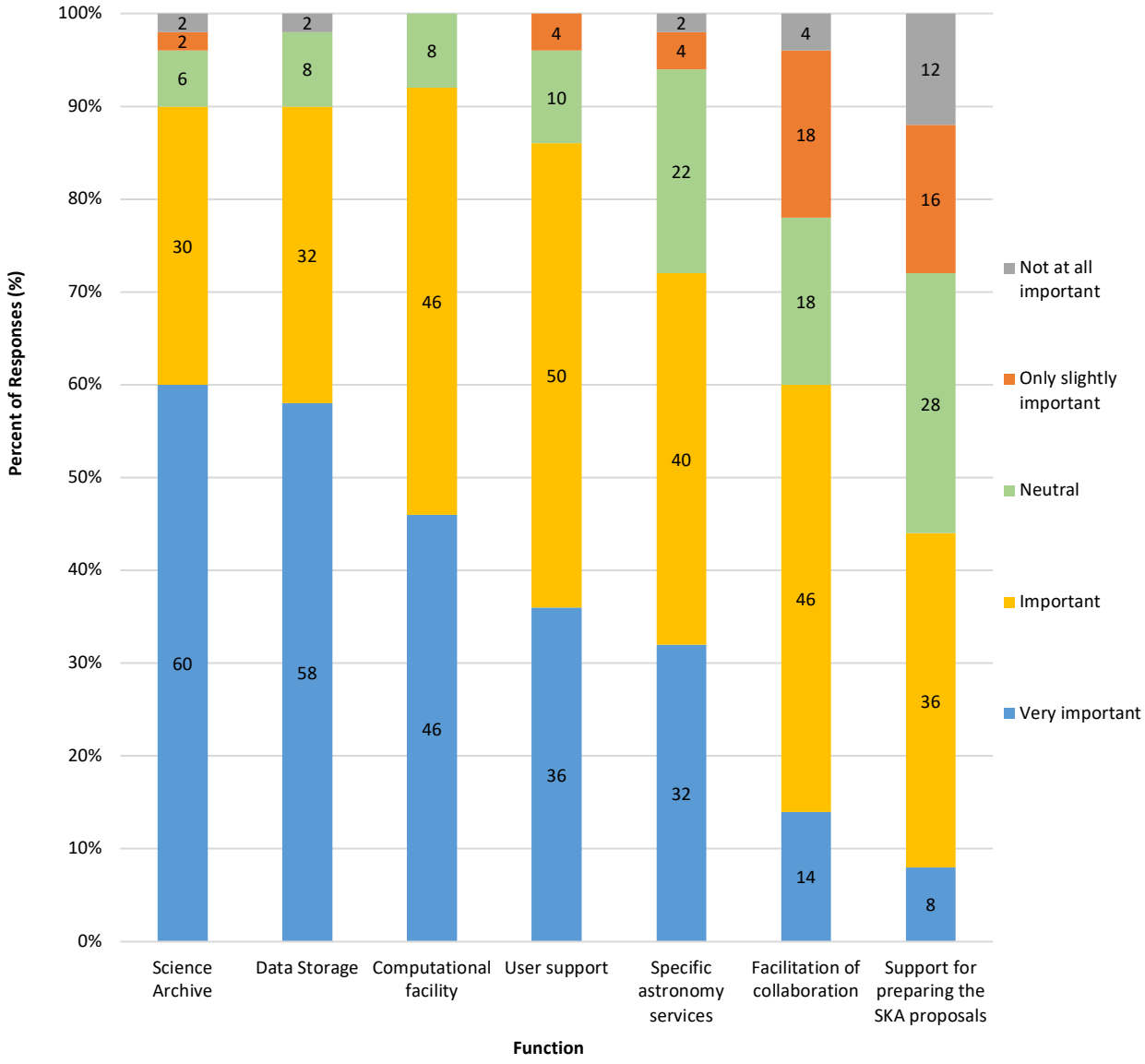
- Low reliability of HPC systems;
- Not being awarded (enough or any) compute time;
- Long time in the queue;
- Issues with HPC filesystem (speed, accessibility of data);
- Inability of current environment to accommodate future science requirements – storage and processing requirements;
- Long processing times (iterative and time consuming to achieve a satisfactory standard during data processing);
- Memory limitations (insufficient memory available in the system).

The graph below shows the distribution of the responses recorded in the list above.



6.6. Expectations of AusSRC

Respondents were asked to rate the importance of a list of functions a future AusSRC should provide.



7. Discussion

7.1. Users and survey confidence

We estimate that the received seventy-four responses constitute approximately 60 to 70 percent of active radio telescope data users.

Responses presented in section 6.1.7 (Do you participate in any SKA related activities?) indicate that almost 63% of the respondents are one way or another involved in SKA related activities.

The responses to questions in sections 6.1.2 (What position do you hold?), 6.1.3 (What project role do you hold?) and 6.1.4 (What is your skill level working with radio astronomy data?) give a high degree of confidence that this survey represents the view of the community highly experienced in working with radio astronomy data.

7.2. Computing skills level

It is somewhat expected that the Python programming language is the most popular choice when processing data, being used by 41% of respondents. C/C++ ranked the second most popular language, used by 20% of respondents (see section 6.1.5 Which programming languages and techniques do you use often?).

75% of the respondents who are using Python estimate their skill level at Intermediate level or above. 50% and 32% of the respondents using C and C++, respectively, estimate their skill level as Intermediate or above.

The important skill of parallelizing the code is only available to 11% of the respondents, and only less than one third of those (29%) feel confident in using it, with those who indicated an expert level, only 7% (see 6.1.6 What is your skill level with these languages and techniques?).

7.3. Projects

71% of the respondents use Australian telescope facilities in their observational projects (see 6.2.3 Which radio astronomy facilities do you currently use?).

78% of the respondents are collaborating internationally (see 6.2.2 What is the nature of your current collaborations?), and 63% are involved in SKA related activities (see 6.1.7 Do you participate in any SKA related activities?).

The range of projects is quite broad with the continuum observations being the most common type of observation (37%), followed by the spectral line observations (30%) (see 6.2.4 What is the observational nature of the data you are using?).

The responses in Section 6.1.1 (What are your areas of scientific interest?) and Section 6.2.4 (What is the observational nature of the data you are using?) demonstrate the broad and uniform distribution of observational radio astronomy in Australia.

7.4. Data

The responses in Section 6.3.1 (What is the size of the data with which you interact?) indicate that the users are already working with data sets of significant size – hundreds of gigabytes, many terabytes. The location of the data is more or less evenly divided between local storage (35%), HPC/Data centers (30%), and institutional data centers(29%). Only 6% of the respondents use public clouds to store their data (see 6.3.2 Where is your data stored?).

Sections 6.3.3 (What are the biggest issues you encounter when working with radio data?) and 6.3.4 (Biggest issues working with the data – elaborated) detail the most common challenges encountered when working with data. They currently represent the impediments for more productive science with the SKA precursors. The scale of the SKA data will only exacerbate this situation, unless suitable infrastructure and software technologies are established.

7.5. Software and interaction with the data

Section 6.4 Software and Tools Profiling of the survey provides a view on what software tools are currently important to the users. The vast majority of those tools are unable to support the SKA scale data sets, and are not providing a mechanism for client-server or remote interactions with the data.

In addition, Section 6.4.4 (What are the biggest issues that you encounter using astronomy software and tools?) indicates that the most common problems with the software are related to: poor quality of code, insufficient documentation, lack of user and software support, lack of support for different platforms, lack of support for the HPC environment.

Despite many users now working with large datasets, 29% of the users are still using desktop/laptop computer to process their data (see 6.5.1 Where do you process your data?). Pawsey and local computational clusters are used by 21% and 19% of the respondents respectively. 16% are using institutional HPC resources.

7.6. Gaps and challenges

In answering 6.2.5 (What is the most challenging stage in observational projects?), 58% of the respondents have indicated that data processing is the most challenging part of their projects. 38% of the respondents also named “data processing” as the most challenging part of working with data in 6.3.3 (What are the biggest issues you encounter when working with radio data?).

In Section 6.5.3 (What is your level of satisfaction with the data and compute facilities you use?) the respondents were asked to estimate their satisfaction with the data and computation facilities that they are using. The levels of satisfaction, vary from 29.2% for the job queuing to 52.9% for the compute speed. The levels of dissatisfaction, however, vary from 17.2% for the user support to 35.3% for the storage capacity. Amongst the largest problems 26% of the respondents are dissatisfied with the reliability of HPC and data

resources, and 26% are not receiving enough or any compute time on HPC. These are important indicators where AusSRC could position itself to improve the experience of users in processing data.

7.7. Expectations from AusSRC

Section 6.6 indicates that significant expectations from AusSRC exist in the community. No proposed areas scored less than 46% (Support for preparing the SKA proposals) as being important and very important. The areas that have scored 90% and above or just slightly below includes science archive, data storage, computational facility, and user support. 72% of the respondents also believe that specific astronomy services are also an important area of development for AusSRC. 60% of the respondents believe that AusSRC should play a role in facilitating collaboration.

8. Conclusion

The survey was conducted with the primary objective of taking a snapshot of the radio astronomy data user community in Australia in the context of the SKA and its precursors to inform the development of the Australian SKA Regional Centre.

Designing the survey, the authors aimed to achieve the following:

- Identify the current gaps in the services, support and infrastructure that is used today to store and process radio astronomy data;
- Map existing needs and expectations in the community onto the requirements from AusSRC and
- Assist with identifying the requirements for the global network of SRCs.

Seventy-four survey responses were received with good coverage of project roles (from project leaders to students), high qualification of data users, as well as the broad and reasonably uniform distribution of research interests covering many fields of observational radio astronomy. It sets confidence in the statistics obtained through the survey.

It was observed that the respondents have good connections to SKA activities, which allowed linkage between the findings and the objectives and requirements for the future Australian SKA Region Centre and its current design study activities.

The survey indicates the current existence of a broad range of types of observational projects, as well as a broad range of telescope engagement, both nationally and internationally.

The survey indicates several gaps and challenges in making science with today's radio data:

- Processing is a significant challenge across all stages of the data path, including, moving and storing large volumes of data, as well as the computational reduction of this data.
- HPC support of existing software is minimal, and maintenance of the software is often poor.

- More than 2/3 of the users are using HPC facilities to reduce data; however, the lack of dependable/dedicated HPC is a significant issue.

The survey reveals the expectation in the community that the SRCs are to be an operational resource for the SKA Archive, Store, and processing needs primarily, but not the individual research centres.

This survey is a snapshot of the current state. It would be reasonable to expect that with the ASKAP surveys now underway, the development of new software tools, and the refreshed infrastructure at Pawsey, the focus and challenges may shift over time. AusSRC Design Study Program needs to track and follow such changes; therefore, we anticipate taking another such a snapshot survey in November 2021.

9. Acknowledgements

The authors would like to thank all the respondents for taking time in answering the survey questions.

We would also like to thank AusSRC Management Committee for supporting this work.

VIII. REFERENCES

[RD1] AusSRC White Paper, 2018, https://aussrc.org/wp-content/uploads/2019/05/AusSRC-White-paper-v1.pdf
[RD2] AusSRC Community Workshop: Summary of Discussions, 2017, https://aussrc.org/wp-content/uploads/2019/05/Australian-SKA-Regional-Centre-Workshop_-Summary-of-Discussions.pdf