SKA Regional Centres

A White Paper by the SKA Regional Centre Steering Committee

V1.0, May 2020

Peter Quinn (Chair, Australia), Michiel van Haarlem (Deputy Chair, Netherlands), Tao An (China), Domingos Barbosa, (Portugal), Rosie Bolton (SKA), Antonio Chrysostomou (SKA), John Conway (Sweden), Séverin Gaudet (Canada), Hans-Rainer Klöckner (Germany), Andrea Possenti (Italy), Simon Ratcliffe (South Africa), Anna Scaife (UK), Lourdes Verdes-Montenegro (Spain), Jean-Pierre Vilotte (France), Yogesh Wadadekar (India)

PURPOSE AND AUDIENCE

This white paper is intended to provide a common descriptive framework and dialogue for SKA Regional Centres (SRCs) that will enable the SKA Observatory (SKAO) and the SRC Steering Committee (SRCSC), in consultation with the SKA community, to jointly define the function and form of an operational SKAO/SRC collaboration that maximizes the on-going scientific outcomes of the SKA. The audience for this paper covers the SKAO Board/Council/CPTF, its advisory bodies and the national efforts aimed at developing individual SRCs including their stakeholders and funding agencies.

OVERVIEW

The resources needed to fully process, distribute, curate and utilize data flowing from the SKA are currently beyond the scope of the SKA1 construction and operations budget. As previously experienced by the Large Hadron Collider (LHC) project, the SKAO and the international SKA science community will need to work collaboratively to shape and establish a shared, distributed data, computing and networking capability that draws on international cooperation and supports the broad spectrum of SKA science. This distributed and shared capability needs to be persistent and needs to utilize a variety of sources of funding from multiple governments and stakeholders. It needs to also be coherent and logistically centralised in terms of the supported services and shared resources that enable a coordinated functionality in support of both the SKA science community and the SKAO. This persistent and coherent set of infrastructure and services can only be achieved through a strong governance model that allows coordination of distributed development and operational effort, as well as a high visibility of the criticality of this enterprise to the major SKA government stakeholders. This White Paper explores these issues, outlines a possible governance model, identifies some of the required core functionalities to support both the SKAO and the science community, and provides an initial estimate of the likely scale of resources and costs involved for a SKA1-class global network of SKA Regional Centres.

1. SRCs – THE NEED AND PATH FORWARD

Since the development of digital detecting systems for astronomy in the 1970s/80s, and the ready availability of inexpensive and deployable "commodity" computing systems at around the same time, astronomy has become a data driven and data intensive science. In the period where astronomical data flows were consistent with commodity computing, it was possible to consider building a new facility that provided end-to-end coverage of the data processing, storage and archiving needs of its community, and where that community broadly had ready access to the resources needed to produce a maximal scientific outcome.

In the current world of large-scale, global, science-driven projects, this approach is no longer feasible, as was first demonstrated by the pioneering CERN LHC more than 15 years ago. These projects now require the incorporation of what was previously considered specialized, rare and centralised High Performance Computing (HPC) and data analytics systems into the very fabric of operational facilities. This shift away from localized small computers and clusters to wide-area-accessible, large-scale facilities is producing a significant change in the whole ecosystem of research communities. Furthermore, the increase in data volumes and processing needs over one research "generation" (factors of 10-100 for the SKA within 20 years) means that traditional data processing methods will most likely break and innovation is needed in algorithms, methodologies and even the basis for the digital representation of data.

Recognizing this reality, the SKAO Data Flow Advisory Panel recommended in March 2016 [1] 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". Following the endorsement of this recommendation by the Board, the SKAO formed the SRC Coordination Group (SRCCG - first meeting 23/9/16), with representatives from SKA member states (represented regionally) with advisors from the Vera C. Rubin Observatory and LHC projects. The SRCCG began the process of defining some of the basic requirements and challenges for a system of SRCs ([2],[3]) and worked with national communities as they sought to create proto-SRC projects. By the end of 2018, proto-SRC design and development projects were in advanced stages of planning and initiation across 13 SKA member states. This evolution from conceptualisation to implementation required a parallel evolution of the SRCCG to a new body that could take this work forward on a global basis. In November 2018, the SKA Board approved the formation of the SRC Steering Committee (SRCSC). The mission of the SRCSC [4] is to:

"Guide the definition and creation of a long-term operational partnership between the SKA Observatory and an ensemble of independently-resourced SKA Regional Centres"

To achieve this mission, the SRCSC has identified the need for a White Paper (this document) as the starting point for the formation of an operational SRC partnership and collective set of SRC services. The collection of both services and infrastructure that constitute a global SRC capability will be referred to in this White Paper as the "SRC network". This paper will not address in detail all of the political, technical, financial and scientific issues that need to be addressed in order to form the SRC network. That work will be undertaken in consultation with SKAO, the Council, its advisory committees, the proto-SRC projects and the broader SKA community over the next 2-3 years, facilitated by SRCSC working groups. What is important for the present is that this White Paper addresses the following items:

- **Governance and operations**: The basic governance and operational model and structures; and the agreements and commitments that need to be created in order for the SRC network building to commence.
- **Baseline functionality of the SRC network**: It is critical that SKAO advisory committees (e.g. SEAC, Operations Review), SKAO operational planning, the SRC infrastructure providers and the broader research community are well connected for co-developing the SRC concepts and co-define the functional scope for a minimal SRC network in the medium and long terms.
- **National commitments**: Finally, it is also critical that SKAO member states and SRC stakeholders understand how they will engage with the SRC network, what the potential scale will be of their national commitments to SRC network capabilities and how their contributions factor into their total SKA investment and the needs of their research communities.

These three components will be the main focus of this paper together with a picture of what an operational SRC network would look like and an outline of the planning needed to progress SRC capabilities over this decade.

2. GOVERNANCE AND OPERATIONS

2.1 Collaboration and Coordination

The group of national proto-SRC projects that are currently underway are independently surveying the future needs of their communities. They are laying the groundwork for planning that will lead to national or regional applications for funds to initiate the design, development, construction and eventual operation of nation-based and/or region-based SRC capabilities. The group of national and regional proto-SRC projects is also now engaging in a joint international effort through the SRCSC to define and shape the collaboration, coordination, interfaces, processes and capabilities that will be needed for the SKA community to address the processing diversity, data logistics and archiving needs of an operational SKA Observatory. While the national or regional SRC efforts may differ greatly in scale, timeline, resources, interests, stakeholders and priorities, each will need to contribute to a converged, coherent, and logistically centralised international SRC network that meets the needs of an operational observatory while being responsive to the needs of key projects, teams and science cases identified by SKAO and of the full cycle of SKA data use.

It is clear that a mechanism for global collaboration and coordination of distributed SRCs will be required. This mechanism must accommodate the likely boundary conditions and requirements that are set by the realities of funding sources, rapidly evolving technologies and the need for a consistent and uniform presentation of SRC functionality to an operational observatory and to a distributed global user community. In particular:

- The SRC network must be persistent. The operational capabilities of the SKA Observatory and its scientific return depend on this persistence. Therefore, any collaboration and coordination mechanism must be capable of being persistent, agile and independent of the diversity of national SRCs with regard to funding timelines and capabilities.
- Any mechanism for collaboration and coordination must be able to function in a landscape where the funding and resourcing of any individual SRC is owned and controlled at the national and/or regional level by diverse sets of funders and stakeholders.
- Any mechanism for collaboration and coordination must ensure coherence, consistency and uniformity of distributed services that can be deployed and supported by and across the SRC facilities in support of both the SKAO and the community. That mechanism must be responsible to the SKAO and the community for this operational consistency and uniformity. That responsibility can only be enacted via a commitment to (and accountability for) the deployment and support at the individual SRC level of consistent core-services defining a distributed services platform at the collective SRC level. The consistency, uniformity and scalability of the SRC distributed services platform that any mechanism is responsible for, can only be achieved through the summed commitments of individual SRCs, coordinated and delivered in the absence of direct financial or managerial control of the individual SRCs.

Clearly the capabilities of the SRC network (as coordinated and delivered by the mechanism in question) are critical to the success of the SKA Project over the lifetime of the SKA. Equally clearly the coordination and collaboration mechanism itself, and the underlying SRC capabilities at national levels, needs to be endorsed and supported through the work of individual governments within the SKAO IGO. Without that support and long-term engagement by government entities within the IGO, the SRC network (and the deliverables it is responsible for) will not be persistent. The collaboration, coordination, and operation mechanism must therefore strike an appropriate and agile balance between control and collaboration, between collective and individual commitments and between logistically centralised and local operations so as to enable persistently high SKA scientific returns.

2.2 The Mechanism: A governance and operations model

A mechanism for collaboration and coordination will be defined through an appropriate governance and operations model that responds to the requirements outlined above. It is the view of the SRCSC that the governance and operations model will involve two distinct layers of responsibility and accountability.

1. A Governance and Management Layer: To meet the needs of the SKA community and SKAO, this layer shall be responsible for the management and development of the persistent capability, uniformity, scalability and performance of the SRC network of services, shared resources and user support, including policies for flexible resource allocation. [SRC Coordination Committee]

2. An Operations Layer: The day-to-day operations of the systems and interfaces between the SKAO and the SRC network that ensure uniformity and service delivery across an agreed set of services and capabilities for SKA science [SRC Operations Group]

This two-layer structure separates and articulates two different levels of concern and responsibility: the overall governance and management, and the explicit day-to-day operations. This is consistent with the governance and operations two-layered structure that was identified and implemented by the LHC project and which underlies the WLCG (Worldwide LHC Computing Grid). It is the opinion of the SRCSC that such a WLCG-like two-layer structure should be adopted as the starting point for an overall governance and operation model for the SRC collective network and its inter-operation with SKAO. In particular, the contributing parties to the WLCG are joined to CERN and the LHC project through a set of MOUs. Similar MOUs would define the commitment each national SRC effort will make to the collective SRC network and represent the mechanism through which individual funding agencies and governments are connected to their community's planned commitments to the SKA with regard to data processing and management needs. The following are potential models for the two SRC bodies introduced above:

SRC Coordination Committee (SCC):

- A committee consisting of national representatives from each member state contributing SRC resources, and representatives of the SKAO
- With a Chair elected from within the committee, the SCC reports, in consultation with the SKAO DG, into the SKAO Council to ensure governmental visibility of SRC commitments and resource management policy
- The SCC will form and manage subcommittees and working groups consisting of staff from the member SRC efforts, and SKAO, that control the necessary resources and personnel for, in particular: co-designing and co-developing the coordinated and shared SRC network of distributed services and capabilities; co-designing and co-implementing flexible, resource management and allocation policies; co-designing and co-implementing the SKA data logistics between SKAO and the SRC network as well as community-facing operational capabilities
- Each SRC represented in the SCC has signed a fixed term, renewable MoU with the SKAO that:
  - Ensures an SRC's compliance with, and support of, SKAO policies and standards
  - Ensures an SRC will make a contribution ("pledge") to the operations and agreed functionality of the collective SRC network over an agreed minimum period covering personnel, computing, storage, data management and communication resources and services (not necessarily the same mix for all parties nor constant in time)
  - Ensures that each SRC will work within the coordination mechanisms (subcommittees) of the SCC to co-develop and support shared SRC network services and capabilities, to support the SRC network efficiency and performance, to co-develop and support flexible, logistically centralised resource management and allocation policies.

SRC Operations Group (SOG):

- A joint operations team consisting of staff contributed from (and employed by) the individual SRCs (and facility providers) plus SKAO operations staff, that is collectively responsible for the day-to-day operations at the interfaces between the Observatory and the SRC network, in particular the services (computing, storage, data management, communications) and shared resources required for the data logistics between the SDPs and the SRC network. Individual SRCs will be responsible for their community facing operational obligations.

- The SOG shall be part of the broader SKAO operations team and will report to the SKAO Director of Operations for SKAO-facing obligations. There are also community-facing operational obligations that will involve SOG team members and other operations staff at the SRCs. This distributed community-facing operational team shall report to the SCC and for the moment (pending a full SRC network operations plan) is represented as a single SOG reporting line. SOG team members may be distributed between the SKAO HQ and individual SRCs.

This two-layer governance and operations structure addresses the persistence, coordination, independence and coherence requirements on the SRC network as described above. Persistence arises through the direct connection and visibility the SCC provides of the national SRC efforts and SKAO pledges to the government representatives on the Council. Governance and coordinated management is the core work of the SCC who are responsible for the coherent shaping, co-development and co-management of the global SRC network services and shared resources. The equal representation of the member states contributing SRC resources and the SKAO around the SCC table, allows the independent status of the SRC national/regional efforts to be recognized within a collaborative forum with the SKAO. The close co-operation on operational processes between the SRCs and SKAO, enabled by the SOG, provides the critical coherence needed for an operational observatory. The proposed model is inspired by the LHC WLCG model, even though the diversity and complexity of SKA science-driven workflows differs from those of the LHC. Nonetheless significant lessons can be learned from LHC/WLCG operations over the past decade.

SRC Network Governance

3. SRC NETWORK FUNCTIONALITY

The overall functionality of the SRC network (infrastructure plus services) will be driven by both the needs of the observatory to deliver data to users and the needs of the user community to deliver scientific outcomes. To support SKAO as it completes the planning for observatory operations and begins the process of ramping-up operations personnel and systems, there must be a clear definition of the minimum functionality that the Observatory can expect from the SRC network. Similarly, the community needs a clear definition of the types of services they should expect from an SRC network and how those services will develop in response to evolving community needs and community provided developments. The “observatory facing” and “community facing” functionality and interfaces will not be the same. The SRCCG (in the past) and SRCSC (currently) have begun the task of defining the baseline functionality of SRCs and the SRC network, and mapping the set of user groups and their varying requirements and associated processes. A detailed analysis is outside the scope of this white paper and will be conducted by SRCSC working groups in the coming 12-24 months. This analysis will draw directly from input provided by SWGs/KSPs and the work of the recently completed AENEAS project that addresses many of the core functionality issues of an SRC network [5]. It will also draw upon practical experience already being gained through proto-SRC efforts, the first operational example of which is the recently reviewed China proto-SRC [6].

At a high level, the community and observatory facing functionality of the SRC network, across all user groups, has six components:

- **Data Logistics** - deliver data from the Observatory to SRCs; make data accessible by authorized users; enable the flow of data between SRCs as required (e.g. for processing, archive workload sharing and data management).
- **Commonality** - support a common and minimum set of tools/interfaces/systems, within appropriate standards, that can be deployed across the SRC network nodes, enabling efficient and collaborative use the SRC network computing, storage and communication resources.
- **Data Processing** – provide the computational and storage resources and services, together with programming and execution environments, needed to interact with and analyse the delivered data. This includes data logistics across and within the SRC network nodes following science-driven processing workflows.
- **Data Archiving and Curation** – provide a functional and persistent SKA Science Archive incorporating FAIR and VO services that allows data discovery, access, use and reuse, new science and scientific reproducibility as well as data provenance all along the full data cycle.
- **Scalable Resource Management and Allocation** - enable and support an interface to the observatory Time Allocation Committee (TAC) and to the broad SRC user community, allowing a variety of science-driven

applications to share efficiently across and within the SRC nodes a flexible set of logistically centralised core services and the resources they require.

- **User Support** - supporting all users with all of the things above

It should be noted that some user groups (such as SKAO/SRC Operations) would access both observatory and community facing functionality. It should also be emphasised that this functionality is the collective and aggregate SRC network functionality and not necessarily the scope of the functionality for any individual SRC. The following tables outline the specific user group functionalities and the minimum targets for the start of SKA full science operations.

3.1 Observatory Facing SRC Network Functionality

User Group	Functionality	Minimum Functionality
SKAO Time Allocation Committee (TAC)	● Provide technical assessments of the SRC resource requirements identified in SKA PI and KSP proposals ● Participate in a coordinated/integrated TAC process to optimally connect pledged SRC resources with high priority science programs	● Have personnel, processes and systems in place for 100% baseline functionality 12 months before 1st TAC covering TAC specific tasks and pledged storage/processing capabilities that can be allocated to successful proposals
SRC Operations Group	● Secure distribution/delivery and reception of released SKA1 data products including data flow management, resource management and reporting ● Adherence to agreed uptime and quality of service metrics ● In the event of data that has PI/KSP identified issues, report/resolve issue and retransmit data to SRC as required	● Ability to receive and manage, 100% of predicted SKA1 initial data flow 12 months before start of full science operations

A definition for the operational interface between SKAO and the SRC network surrounding the TAC process is still to be discussed and agreed and forms part of the SRCSC work program (section 8). A preliminary description of the SKAO/SRC data transmission and data management interface has been defined [7] and will evolve as part of the SRCSC work program.

3.2 Community Facing SRC Network Functionality

User group	Functionality	Minimal Functionality
SKA Programme users	These can be users at the stage of Proposal preparation, or (astronomy) science users either in the Project Execution stage or in the Science extraction stage. In the last two stages they would have active projects with awarded time in the SKA Observatory - either as members of a Key Science Project or as Principal- or Co- investigators on smaller projects. These users will have authorised access to proprietary data located in the SRC network and authorised access to processing and storage for their project needs allocated as part of the proposal review process, and in line with resource estimates taken into account when they were awarded time.	• 100% support for 1st proposal cycle in terms of data distribution, data access and data processing at the TAC agreed levels in line with the baseline data storage and processing capacity and TBD common/supported processing toolsets and workflows • Functional SRC network gateway for SKA Program users covering data access, logistics and processing workflows • Baseline user support personnel in place
SKA Science Archive users	These are (astronomy) science users typically accessing public data in the SKA Science Archive to undertake their investigations. The projects they work on do not require additional SKA Observatory time, but rely on access to data and data products already existing in the Science Archive as well as to provenance information that allows their reuse. Depending on the scale of resources needed, these users may need to request access to processing and storage for their project needs but the process of allocation of these resources in the SRC network is TBD. The Science Archive is a unified facility consisting of a common portal supported by distributed SRC resources.	• 100% support for agreed baseline SKA Science Archive Functions (TBD) within 12 months of completion of 1st proposal cycle • Functional SKA Archive portal covering data access, data discovery, interoperability and processing workflows covering 100% of baseline SKA Science Archive functionality • Baseline user support personnel in place
Non-astronomy users	These are users of SRC network resources who are undertaking work that is neither astronomy related nor typically using data products in the SKA science archive. (Examples could include researchers from Open Science disciplines looking at how data are used, or representatives from funding agencies looking at usage data.) This is expected to be a very small fraction of the real work of SRCs	• No guaranteed functionality by operations start
Software developers	These are software developers who make use of the distributed computing tools in the SRC network for testing that their software can be run effectively in the SRC network environment.	• TBD Baseline support by 2025
Tools users	These are SRC users who wish to access software tools made/developed for, or used by,	• No guaranteed functionality by

	the SRC, but for use on external systems - recognising the fact that the SRCs will serve as excellent repositories for software, workflows as well as for data.	operations start

A full definition of the baseline functionality (data access, data processing, workflow design, repurpose/reuse and execution) for the SRC network Science Gateway (including the set of supported observing modes) will be developed jointly by the proto-SRC projects and the SKAO within the SRCSC (and SCC) working groups (that include community user group representation). A similar process will be involved with the baseline functionality definition of the SKA Science Archive.

4. SRC COMMITMENTS

The design, development, operations and maintenance of a global network of interoperating SRCs that meets both national/regional and SKAO needs and priorities, is going to be a complex, costly and critical task. While the LHC has successfully built and run a similar system, and there is a lot to be learned from their collective experience, the diversity of science-driven large-scale workflows in the astronomy community raises new challenges. Astronomy has acquired considerable expertise in distributed data systems, automated processing and interoperability over the past twenty years through projects like HST, VLT and ALMA and the work of the International Virtual Observatory Alliance (IVOA). Similarly, accessing global data and computing resources has become commonplace through the development of data content management systems (Google, Facebook, the WWW in general) and cloud computing providers. This was not the case for the LHC who had to invent “the Grid” to fulfil much of the functionality we now take for granted. We should therefore be somewhat optimistic that the SKA community (learning from the LHC experience and working with infrastructure providers and modern data/cloud systems) can build and run a capable SRC network. We will also need to crawl before we walk. Thankfully we have a number of now operational SKA precursors and pathfinders, (MeerKAT, MWA, ASKAP, LOFAR, FAST, CHIME, eMerlin and others) that provide an excellent training ground in which to develop SRC functionality and to support the community with delivering SKA precursor and pathfinder science within the coming five years. A nation’s commitment to SRC developments should therefore not be seen as something for the end of this decade when SKA1 is operational but rather a commitment that starts now with proto-SRC projects that connect to SKA precursor needs. At this time there is already a collective contribution of over 25 million Euro (more than 50 FTE) from 13 SKA member states and the SKAO, for proto-SRC projects over the next 3-5 years. These proto-SRC projects will provide the resources necessary for the early development of SRC network prototypes.

What an individual SRC will do in the era of an operational SKA1, and what commitment it requires from its community, will depend on:

1. What is finally agreed to be “outside the fence” of the SKA Project in terms of processing and storage capabilities for SKA science data. We have a reasonably clear estimate of this
2. What the scope of SKAO Operations will be – will SKAO be responsible for all user support and quality control of data across the “end-to-end” process, and at what scale? This needs to be clarified particularly in the context of an SKA science archive capability
3. What is the interest and capacity of a national community in terms of SKA science and how does that interest and capacity couple to the needs of other national priorities that might need SRC-like capabilities?

The following sections look at each of these questions in turn to give an impression of likely SRC network costs and individual national commitments.

4.1 Scale and Costs

4.1.1 Storage, Processing and Network

Studies of the aggregate requirements for processing and storage within the SRC network have been conducted by the SRCCG [3], the AENEAS project [8] and the SRCSC. These studies give broadly consistent results for the data flow rate an SRC network would need to absorb and store for the long-term, and the data logistics and processing power needed to convert the Observatory Data Products into Advanced Data Products by SKA Principal Investigator (PI) and Key Science Project (KSP) teams. These studies do anticipate some of the diverse needs of different KSPs and PIs. While there is some level of consistency of the studies, there are a number of evolving and significant questions that will need to be resolved before more robust estimates of data rate, storage and processing costs can be given.

1. What is the efficiency of radio astronomy computational methods and algorithms when mapped onto state-of-the-art distributed HPC/Cloud systems with hybrid technologies? At this time we believe 10-15% (maximum) of the quoted peak performance of HPC systems can be achieved for current real-world radio astronomy applications
2. What will be the final data storage volume as compared to the data flow output rate of the telescopes and the end products of SKA science data processing? The use of averaging and the development of new compression technologies will all have possibly significant positive and negative impacts on the final storage costs
3. To what extent will cloud computing and technologies be relevant to the operations of the SRC network? There are several studies [9] that already indicate radio astronomy workflows at the precursor level can be successfully conducted in hybrid (cloud + monolithic HPC) systems. The capital costs for HPC hardware purchases and refresh cycles, plus continuous operational costs for HPC centres, can be converted into a single, and on demand, cost for broadly accessible cloud resources
4. What is the availability at a national or regional level of HPC-class research infrastructure and academic networking? The open availability ("free") or not, of computational and data-intensive national/regional research capabilities will potentially have a significant impact on the cost basis (capital and operations) for any given SRC

Bearing in mind these open issues, the studies to date indicate the aggregate levels of bytes (storage of one copy of all data products) and flops for an operational SKA1 SRC network shown in the table below. Additionally, detailed studies by the AENEAS Project [10] of the international networking required by these data volumes around an SRC network indicated a mean speeds of 100 Gb/s.

Data Flow PB/yr	Processing PFlop/s	Network mean speed Gb/s
710	22	100

4.1.2 Personnel

The provision of personnel to design and build the SRC network, to run SRCs and provide services to their user communities, and to contribute to SKAO operations, will be a significant cost for the SRC network above that associated with hardware and networks. Given that discussions on the scope of SKAO "inside the fence" operations are not yet complete, and given that we are still at an early stage in discussions on the range of services and capacities the SRC network will provide, there has not yet been a detailed study of the personnel costs for SRCs. Within the SRCSC, there is considerable experience based on the development and operations of distributed data systems (e.g. HST, Gemini by CADC), distributed data operations (e.g. VLT and ALMA at ESO) and distributed international user support (ALMA regional centres). Based on these insights, the SRCSC believes a reasonable estimate of an operational SRC network at SKA1 scale will require an aggregate operational/maintenance workforce of order 100 FTE.

4.1.3 Aggregate Costs

Based on the studies and estimates mentioned in the previous sections the likely annual costs of storage (1 distributed copy), processing (assuming a 5 year buy/refresh cycle), networking and operations/maintenance personnel are summarized in the following table.

Data (M€/yr)	Processing (M€/yr)	Network (M€/yr)	Personnel (M€/yr)
18	2.4	5	10

This equates to an annual full SKA1-scale cost of **35.4 M€/yr.** Allowing for the uncertainties outlined in the previous sections, the likely cost range is **20-40 M€/yr.** Over the coming ten years, it is assumed that the personnel costs will shift from development to operations/maintenance, maintaining an aggregate of order 100 FTE.

4.2 Participation

The sizes and numbers of SRCs around the world, and the topology and capacity of the SRC network, will be largely determined by the willingness of SKA member states to participate in the SRC network and what form that participation will take. It is likely for purely practical reasons that SRC capabilities will be developed close to the observatory processing centres in Australia and South Africa. There is already a keen interest from some non-host countries (e.g. Canada, China) to develop significant SRC capabilities and for some regions (e.g. Europe via a continuation of the AENEAS collaboration in the framework of the EOSC and EuroHPC initiatives) to develop collective/coordinated SRC capabilities. It is also clear that the business cases individual countries need to develop with their funding agencies to support national SRC projects, are not all the same. In some cases, SRC funding will come as part of a bigger national agenda covering data and computationally intensive research in many fields. In other cases it will be seen as part of a direct and connected contribution to the SKA project along with construction and operations funding. This implies the government departments, funding agencies, stakeholder communities and KPIs underlying national SRC efforts are likely to be different, as will be the motivation for SRC participation. Given this diversity, the SRCSC has tried to focus on two central issues in order to inform SKA Council thinking on SRC participation.

1. What is the baseline participation needed to ensure end-to-end SKA science opportunities in the SKA member states? The costs outlined in section 4.1 represents the baseline aggregate capability needed. The SRCSC proposes that, at a minimum, SKA member states use their fractional contribution to SKA construction and operations as representing their fractional contribution to the baseline operations and development of the SRC network. This contribution could take several forms and would be identified in the MOU signed by a member state with SKAO to formally recognize their SRC effort as part of the SRC network. The associated resources will be contributed (pledged) to SKAO operations and SRC coordination/development under the governance and operations model described in section 2.2. It may also be the case that a particular country wishes to put more resources than its nominal SKA construction-and-operations fraction into SRC capabilities. The contribution of these additional resources may be motivated by particular national SKA science interests, multi-messenger astronomy interests or other national priorities. In this case, any additional resources above the nominal participation fraction would not be covered in the SKAO MoU and would not be contributed resources to the operations and development of the SRC network, unless otherwise agreed. Additionally, it would be the expectation that all member states do contribute to the SRC network at their nominal fractional level. This may be as an individual nation with specific national developments, through a national participation in a regional SRC effort, in collaboration with one or more international SRC network partners or as a direct contribution to the SRC operations group (SOG) in one form or another (to be approved by SKAO and SSC). Similarly SKA member states could partner with non-member states to contribute to the SRC network or non-member states individually may also see merit in contributing to the scientific capability of a network of SRCs. Finally we need to consider the situation in which a member state chooses not to contribute to the SRC network. This is a matter for consideration by the SKAO Council and would potentially entail a restriction of national access to the Advanced Data Products and/or services of the SRC network. These restrictions would not include access to the data provided to the SRCs by the Observatory.

2. Why should a member state participate in the SRC network? The success of the SKA, and the achievement of a maximal scientific return on national investments, clearly depends on both the performance of the telescopes and associated systems, and the science-driven capacity and capability of the SRC network. Without the two working effectively together as one integrated "end-to-end" science system, we will not realize the full potential of the SKA nor the benefits that arise through our communities

developing skills and systems to deal with the modern challenges of massive data and Internet of Things (IoT) technologies. Beyond this "obvious" reasoning, there is also the question of how should/could a national contribution to SRC network capabilities be recognized. The SRCSC understands that this is a complex and important matter that needs to be considered by the SKAO Council in the near term. The SRCSC suggests that, if a return is to be identified with SRC contributions, then it could take several forms including the offsetting of some national SKAO construction and operations costs, the allocation of SKA development funding, additional support from SKAO resources to national efforts (e.g. SKAO staff support for a range of national activities) or SKAO funded fellowships/scholarships/positions at national institutions to support SRC developments. Finally there will be important operational advantages for a member state to be a member of the SRC network. Successful proposals for telescope time will be those that contain both excellent science and the SRC resources needed to deliver the outcomes. The SRC network will be actively involved with SKAO, within the TAC process, in assessing and allocating resources to identify and ensure successful science proposals. Additionally, the SRC network will be a communal science analysis platform that will enable collaborations and new initiatives built on persistent capabilities at the SKA-scale.

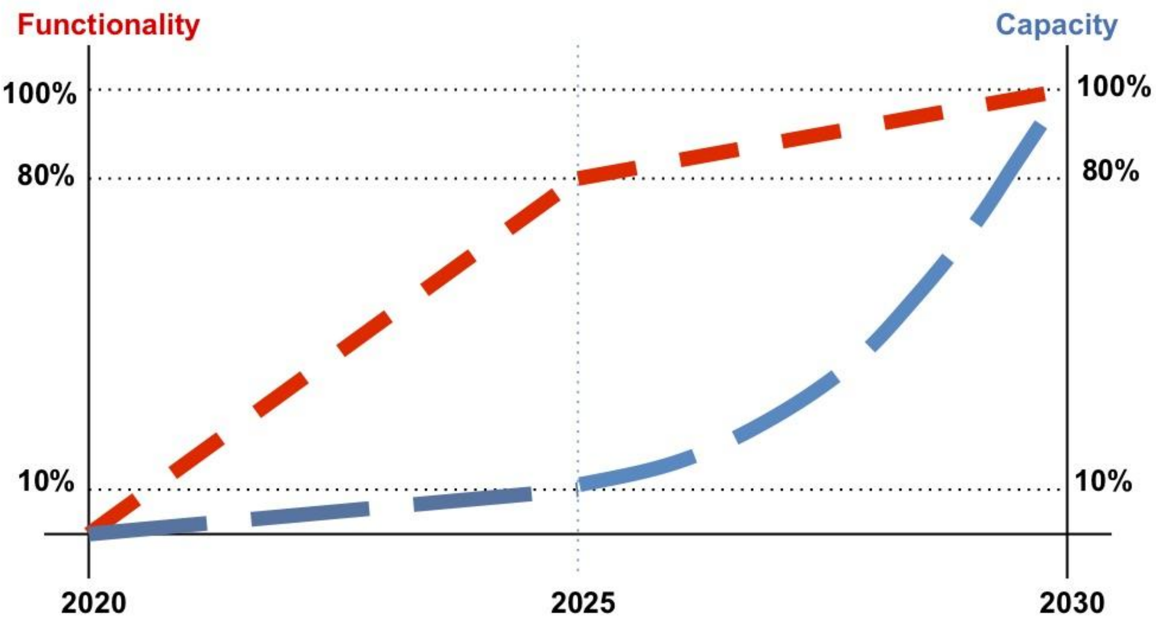
4.3 Road Map

As noted above, we have an ideal opportunity to grow SRC capabilities and skills, and deploy an initial version of the SRC network, through engagement with SKA precursor and pathfinder science programs and early proto-SRC initiatives at national levels. This will bring benefits to the precursor/pathfinder science teams and advances the design of the SKA1 SRC network through real world data challenges and prototypes. These early developments will prepare the proto-SRC network for a possible engagement in early SKA1 science verification/operations. The SRCSC believes it is appropriate to ramp-up the SRC network over this decade in a two-step process.

- 2020-2025
  - Securing proto-SRC development funding and initial engagement in national precursor efforts (many are already at this point)
  - Completion and submission of business cases for SKA precursor-scale SRC capability and participation in SRC network development
  - Through SRCSC (SCC) working groups and subcommittees, and in response to SRC data/processing challenges, develop the SRC network, a common basis for SRC network tools and services and support specific precursor science programs
  - Through a significant allocation/secondment of developers to SCC committees, achieve 80% of SRC network functionality and 10% of SKA1 baseline capacity by 2025
- 2025-2030
  - Secure SKA1 class SRC network funding
  - Continue development of baseline functionality to 100%

- Support early science verification and operations of SKA1 as and when required by SKAO
- Procure processing, data and networking capacity for 100% SKA1 baseline requirement
- Be operationally ready 12 months in advance of the first round of SKA1 PI and KSP proposals

Given that roughly 1/3 of the baseline SRC cost is personnel and 2/3 is storage/processing infrastructure (section 4.1.3), and equating personnel primarily with “functionality” and storage/processing with “capacity”, then any national commitment would be approximately 1/3 of its SKA1 baseline level commitments by 2025 and 100% by 2030. For a **10% participant** in the SRC network capability, this would equate to **0.7 – 1.3 M€/yr by 2025** and **2 – 4 M€/yr by 2030** following the functionality/capacity ramp-up outlined above and shown in the plot below.



Ramp-up of the SRC network

5. AN OPERATIONAL SRC NETWORK

The operational and governance structures described earlier show that the existence of a joint operational team of staff at SKA and within SRCs is anticipated to coordinate the day-to-day functioning of the network of SRCs. There are several areas of management needed, on various timescales:

Annually and on-going

- Pledging SRC resources into the SKA-facing roles and for user-facing purposes
- Roadmap for development of SRC-related technologies, and pledging of (human) resources to enable this – for example, SRC science gateway and collaboration tools, data management and transfer tools, improving best practice
- Development and maintenance of long-term roadmap for evolving the data transfer, data storage and data processing capabilities of the SRC network.

Quarterly

- Determining areas of software development work in the near-term and prioritising backlog items.
- Track performance of SRCs against pledges, both for “SKA-facing” and “user-facing” activities
- Track SKA’s use of pledged SRC resources against the predicted science programme allocation to check consistency
- Track performance of the data transfer network links – are they performing well enough?
- Review of data management policies to ensure appropriate quality of service (trading off performance and cost / capacity needs)

Weekly / Day-to-day: ensure the continuous availability of the SRC network to support the SKA-facing roles and user-facing roles.

- Monitor ability of each site storage element to accept data from SKA sites - both on a medium term timeline when telescope planning occurs and when data is scheduled to be pushed from the SDP sites within SKA to identified SRC sites
- Monitor ability of appropriate sites to accept batch processing jobs
- Monitor ability of appropriate sites to provide interactive sessions for users

- Monitor and maintain network link availability and performance (e.g. through continuous monitoring of links in use and liaison with network providers)

- Apply data management policies (through use of a data management service) to apply agreed rules that improve access to high-demand data products and minimise unnecessary additional copies of low-demand data products, giving better overall performance across the SRCs.

Additionally to the above activities it will be the responsibility of the SKA to manage delivery of Observatory data products to specific SRC sites. It will be the responsibility of each SRC to perform its own internal operational procedures to ensure its availability to the SRC network is in agreement with the terms and conditions of its MoU with the SKA, as well as the meeting the expectations of its national/regional stakeholders.

6. OPEN ISSUES

The primary focus of this paper has been a high level, preliminary look at the Governance, Functionality and Participation aspects of the SRC network. More work remains to be done on fleshing out these three central issues as part of the SRCSC managed core work program (see sections 7 and 8). This paper does not address all of the design and implementation issues surrounding the construction of a SRC network. Many of these tasks will be the primary focus of the SRCSC working groups in the next two years. Through the work of the SRCCG and SRCSC, through consultations with all national SRC efforts, and with comments and advice from the SKAO Board, SEAC and the SKAO Operations Review Panel, a number of open issues have been identified. Some of these are focused on the design and development of the SRC network and some are policy or principles related. The following summarises the main issues and tasks discussed to date.

6.1 Risk

- The SKAO will be a complex operational system of one observatory, two telescopes, three sites and a collective of SRCs. This complexity carries risks (technical, operational, financial, managerial) and there needs to be a full risk assessment covering contingency, mitigation, recovery and redundancy with regard to SRC network functionality as part of SKAO end-to-end operations and community expectations.
- Distributing critical SKA functionality around an SRC network, and having that network funded by diverse national commitments that must sum to a persistent level of service, presents a significant risk to the ultimate scientific success of the SKA. The effectiveness of the relationships that form between the SCC and Council and between the SRCs and their funders, is critical to the SKA. A failure of those relationships to provide persistence and capacity in the SRC network, should be considered one of the highest level risks for the SKA that needs to be addressed in the short-term.
- The SKA Science Archive, a key driver of SKA success, will be designed and deployed as a distributed, but logically-centralized, system across the SRC network. While this (along with other SRC capabilities) entails risk due to the financial underpinnings of the SRCs, the astronomical community is fortunate in having significant experience in designing and operating distributed, multi-wavelength and multi-messenger archive systems.

6.2 SKA Commissioning and Operations

- SRC “pledged” resources - There must be an integrated resource allocation process, considering SKAO resources alongside SRC resources, through a collaboration and coordination surrounding the resource allocation process
- What is expected, and when, from the SRC network, as SKA1 moves

through commissioning, science verification and early operations?

- What, in detail, is the expected output from the SDP systems for various kinds of programs and how will user support be integrated across SKAO and SRC community facing functions?
- Interface Control Documents (ICDs) for the SKAO-SRC and SRC-User interfaces must be defined early in the design of an SRC network architecture

6.3 Implementation

- SKAO and the ensemble of SRCs will need to enter into a co-design process for the SRC network with NREN/Cyber infrastructure providers – how will this occur?
- There will be a minimum set of software systems enabling a common SRC software and services platform – at what level is this commonality above the inevitability of systems unique to nationally funded efforts?
- How will SW licenses for user-generated SW, or IP in general, be managed in the SRC network?
- Who will be guiding the development of workflows to go from SDP data products to advanced data products? A community-KSP-SRC collaboration?

6.4 Policy/Principles

- Does each country have an obligation to provide resources to an SRC network?
- SKAO data placement in the SRC network - this should in principle be driven by optimizing science outcomes, not by political or financial considerations
- What are the underlying policies for the Science Archive covering such things as public data discovery and exploitation? Open science, open skies? What about access to significant Science Archive processing resources? Coordinated or cross-SRC archive research proposals? Ingesting user generated data products into the archive?
- FAIR (Findable, Accessible, Interoperable, Reusable) principles for data and methods should be adopted by SKAO and the SRC network
- SRC contributions to the collective SRC network pool should be codified in terms of capability and not in terms of cost
- IVOA (International Virtual Observatory Alliance) - SRC data and services must be IVOA compliant

7 SRCSC WORKING GROUPS

The SRCSC was constituted to contain national representatives who would be able to allocate resources within, and direct, national proto-SRC efforts. This is essential to begin the process of coordinating an overall proto-SRC network development and to achieve the commonality and consistency the SRC network and individual SRCs will ultimately require. The SRCSC will, through the leadership of its national representatives, establish a set of working groups to define and undertake the detailed work programs needed in the coming two years, and beyond. The following working groups are currently being formed:

1. **Data logistics** - Across the network of SRCs and the SKA SDP sites, the logistics of moving data and managing the locations of copies of data products is a key concern. The overall goal of activities within this working group is to ensure that by the time of SKA performing the first science verification observations, there is a functioning network of SRCs up and running, able to share the burden of storing and serving SKA data products to users. This means that we must have a functioning collection of long-haul network links working at appropriate transfer speeds, connecting data centres with storage elements at sufficient scale enabled as part of a federated data management service

2. **Operations** - The SRC Operations Group (SOG) will be a joint operations team consisting of staff from the SKA Observatory and staff contributed (and employed) by the individual SRCs in the SRC network. Collectively, this group will be responsible for the day-to-day operations of the SRC network, its systems and processes, needed to maintain overall SKA operations between the Observatory and the SRC network. The SOG will be a part of the broader SKA Operations team and will be led by the SKA and thus report up to the Director of Operations for the SKA Observatory. The team will be distributed between the SRCs in the network (and constituent nodes) and the SKA. This working group will consider the following:
   - Ops functional diagram/relation between SKA and SRC operations groups, i.e. the SOG
   - Definition of the quality of service metrics and other KPIs to measure performance against
   - Define the role of the SOG in the allocation process (e.g. technical assessment of proposals)
   - Define how pledges will be requested, managed and monitored
     - What will be the process for pledging resources?
   - Define how the data transfer links will be monitored and what the process will be to follow up when performance drops below expectations (or the QoS defined above) including disaster recovery
   - Levels of user support

3. **Processing and Work Flow**

4. **SKA Science Archive** - This working group will cover the requirements

and design of services, data models, processes and interoperability for a distributed FAIR science archive implementing IVOA standards and protocols based on the collective experience and knowledge of the community. Topics will include:

- Archive User interfaces and Services
- User Support services (different than operational functions)
- User Authentication and Authorization (AAI)
- Data models
- Advanced data products ingestion/archiving/processes/standards
- Workflow/analysis scripts
- Enable the generation of statistics of use and metrics
- Integration with parts of the other SRC components

5. **Compute Platforms** – The compute working group encompasses two distinct themes related to the underlying hardware platform needed to support the mission of the global network of SRCs. The first is to identify, and validate, the set of requirements that drive the functional and performance needs of the platform itself, and the second is to design a system that meets that set of requirements. These two tasks are fundamentally linked by the reality of a limited budget, and will require an agile, iterative approach to minimising the cost-function whilst delivering against the base requirement set. Thus, housing these somewhat disparate tasks in a single working group makes sense.

6. **Science User Engagement** – The final aims of this long-living WG are ***(i)*** to engage the science community in tuning the requirements for the SRC network in order to maximize the science return and ***(ii)*** to guide the scientific community at large towards the new end-to-end procedures that the SKA era will require for performing a scientific program. Topics to be covered include:

- Collect feedback from the reference Science User communities – SKA program users and SKA science archive users – to help identifying a complete set of requirements for, and the implied scale of, the SRC network, and eventually optimize the final list emerging from the exercise above, by e.g. recognizing commonalities.
- Support, stimulate and coordinate a constant interaction among the software developers and the Science User communities to test the progressive implementation of the needed solutions in the SRC network. In particular, to identify cases-of-study based on real and/or simulated data in order to achieve the delivering of some science results since the early stage of a proto-SRC network.
- Coordinate the set up of the "SRC capability challenges", and to accompany the community in the change of mentality required by the implementation of the SKA1 data extraction workflow. In practice, any activity aimed to tutoring and involving the interested community at large to the use of the SRC capabilities, e.g. from planning of the proposals, to performing data reduction and achieving a full

exploitation of the collected data (maybe partially sharing this duty with other WGs).

Each working group will be co-chaired or coordinated by a member of the SRCSC to provide effective SRCSC oversight. The members of the working groups will be drawn from the SRCSC, the proto-SRC projects and their associate user communities, and the SKAO. Each working group will in turn report to the SRCSC on a regular basis and will have a well-defined terms of reference that will be supported by resources provided by the individual proto-SRC projects and the SKAO. Some aspects of SAFe methodology may be relevant to coordinating proto-SRC network developments among the SRCSC working groups.

8 SRCSC PLANNING

The mission of the SRCSC is to "guide the definition and creation of a long-term operational partnership between the SKA Observatory and an ensemble of independently-resourced SKA Regional Centres". At some point, an agreed governance and operational model of the type discussed in section 2.2 will be established and the SRCSC will have achieved this mission. Until that time, the SRCSC will take a structured project and systems approach to the definition and development of the SRC network. That approach will communicate and coordinate with the agile systems approach that is building the SKA. Over the next year, the SRCSC will focus on refining and delivering detailed papers on Governance, Operations, Policies and Risks for discussions with the SKAO and national stakeholders, with recommendations going to the SKA Council. The individual working groups defined above (and more to come) will develop most of the detailed planning documents required to move from the SRC network concepts outlined in this White Paper to a complete set of requirements across all functional and operational levels. Those requirements will be combined into a SRC Level 1 Requirements document. Following agreement on the Level 1 Requirements at national and SKAO Council level, an SRC Architecture and SRC subsystems design set of documentation will be produced through the combined effort of the SRCSC and its working groups. At that stage, the SRC design would be available for external review. Based on that review and refinement of specifications, a full SRC network implementation plan, supported by commitment from the SKA members, will be submitted for Council approval.

Part of the SRCSC working group tasks will be prototyping and technology assessments that lead to a set of subsystems specifications and designs. This work will rely on a close connection to existing SKA precursors and pathfinders to undertake prototyping and a range of SRC network challenges. Depending on available funds across all working groups, the SRC network will be rolled out, building on prototype systems, to meet SKA1 operational (and potential commissioning needs) by the end of this decade. Again, depending on the resourcing levels for SRC development projects, it is hoped that as the SRC network is rolled out, it will prove to be a significant enhancement to the science capabilities and outcomes of SKA precursors and pathfinders.

Over the 2020-2022 period the on-going activities of the SRCSC and its members focus on:

- Support development of national business cases
- Facilitate international & national SRC meetings/workshops
- Coordinate national developments
- Progress SRC network establishment
- Regular reporting to SEAC and SKA Board/Council
- Prototype SRC network functions (precursor & pathfinder focus, SRC data challenges)

The specific SRCSC milestones in the coming 30 months include:

- 2020 Q2 - Present White Paper to Board

- 2020 Q2 - Set up Working Groups - draft Terms of Reference, recruit members
- 2020 Q3 - Complete Terms of Reference for all Working Groups - first round of meetings
- 2020 Q3-4 - First Pass development of Level 1 SRC Requirements (by Working Groups)
- 2020 Q4 - Governance, Operations, Policy and Risk paper (delivered by SRCSC)
- 2021 Q1-Q2 - Working Groups and SRCSC complete Level 1 SRC Requirements
- 2021 Q2-Q3 - Complete SRC Network Architecture
- 2021 Q4 - Review of Level 1 SRC Requirements and Architecture
- 2022 Q1-Q2 - Finalize SRC Subsystem designs
- 2022 Q2-Q3 - Produce and deliver draft SRC Implementation Plan
- 2022 Q4 - Transition to new Governance structure (SRCSC to SCC)

9. CONCLUSIONS

This White Paper is a necessary first step in the definition of an operational partnership, between the SKAO and a set of independently funded SRCs to deliver a coherent and persistent SRC network that matches the science requirements of SKA1. The three cornerstone elements of that partnership are:

- **Governance and operations**: We have identified a two-tier structure similar to the WLCG governance and operations model as being appropriate for the SRC network partnership. This consists of two bodies (an SRC Coordination Committee and a day-to-day SRC Operations Group) that deliver the planning and persistence needed (through direct connections to the SKAO Council and individual SRCs) as well as operational consistency and performance (through embedding SRC staff within SKAO operations). The commitments of individual SRCs, and their compliance with the necessary standards for SKA data, will be defined through individual SRC agreements with the SKAO in a manner similar to the MOU structures of the WLCG.
- **Baseline functionality of the SRC network**: It is critical that SKAO operations and the broader research community are both well connected to the development of SRC concepts and the definition of the SRC network functionality. That functionality will cover data flow, data processing, data curation, user support, a common layer of tools and systems, and overall resource management on a global scale. The minimal SRC network capabilities for each of these functions, in the medium and long terms, will be defined and developed through specific SRC working groups that are managed through the SRC Coordination Committee. These working groups will lean on the work of existing proto-SRC efforts as well as the lessons learned from direct engagement with SKA precursor and pathfinder projects.
- **National commitments**: It is also critical that SKAO member states understand what will be the potential scale of their national commitments to SRC network capabilities. Although there are still considerable uncertainties associated with the costs of storage, processing and networking for SKA1 on a timescale of 5-10 years, several different approaches suggest that a SKA1 capable set of SRCs will collectively cost 20-40 MEuro per year which includes an annual personnel estimate of 100 FTE. We propose that this SKA1 class capability would be rolled out over the 2020-2030 period, achieving 10% capacity and 80% functionality by 2025. A possible model for the scale of national participation is to use the same fractional engagement as associated with construction and operations. Depending on the number of SRCs, this may vary from nation to nation, reflecting national priorities and the availability of external or internal research infrastructures. The form (e.g. services, people, resources) of participation may vary from nation to nation but all SKA member states would be expected to participate in the SRC network through a variety of mechanisms.

The SRCSC has proposed a plan of work that will address many of the detailed issues raised in this White Paper over the coming 24 months in close collaboration with Council, SKAO, its advisory bodies and the broader community.

10. REFERENCES

[1] **Data Flow Advisory Panel Report to the SKA Board**, 17 March 2016

[2] **SKA Regional Centre Requirements**, 23 January 2019, SKA-TEL-SKO-0000735

[3] **A Model for SKA Regional Centre Global Size Estimation**, 3 December 2018, SKA-TEL-SKO-0001000

[4] **SKA Regional Centre Steering Committee (SRCSC) Terms of Reference**, 12 November 2018

[5] **A proposed framework for designing and implementing a Service Portfolio for the European Science Data Centre and SKA**, 30 November 2018, AENEAS GA DELIVERABLE: D6.2

[6] **SKA Data Take Centre Stage in China**, November 2019, Nature Astronomy, Vol 3, 1030

[7] **SKAO to SRC Technical Interface Description**, 26 September 2019, SKA-TEL-SKO-0001072

[8] **Analysis of the computer load, data transfer and data storage anticipated as required for SKA key science**. 30 September 2018, AENEAS DELIVERABLE 3.1

[9] **Imaging SKA Data in Three Different Computing Environments**, 29 December 2015, Astronomy and Computing, Volume 14, 8-22

[10] **Architecture and Cost Model for a World-wide Network for SKA**, 12 October 2019, AENEAS GA DELIVERABLE: D4.4

All reference documents can be found at: https://www.dropbox.com/sh/2isurcn9idf5uby/AADC2BQLHENjASZIa2tEEo6Ka?dl=0