SKAO TO SRC TECHNICAL INTERFACE DESCRIPTION

Document Number	SKA-TEL-SKO-0001072
Document Type	ICD
Revision	02
Author	M. Dolensky, M. Ashdown, R. Bolton, R. Simmonds
Date	2020-03-09
Document Classification	UNRESTRICTED
Status	Released

Name	Designation	Affiliation	Signature	
Authored by:				
Markus Dolensky	ICRAR Technical Lead	ICRAR	Markus Dolensky	
			Date:	2020-03-10
Owned by:				
Marco Caiazzo	Senior System Engineer	SKAO	MC	
			Date:	2020-03-10
Approved by:				
Juande Santander-Vela	SW Lead Product Manager	SKAO		
			Date:	2020-03-09
Released by:				
Joe McMullin	Project Engineer	SKAO	J P McMullin	
			Date:	2020-03-29

DOCUMENT HISTORY

Revision	Date Of Issue	Engineering Change Number	Comments
A	2019-07-26	-	First draft release for internal review and SCCG
01	2019-09-26	-	Revision 01 for signature.
02	TBD		Post S-CDR revision: added appendix for feedback; added SRCnet to List of Abbreviations; Typos in Section 4.3.1 fixed.

DOCUMENT SOFTWARE

	Package	Version	Filename
Word processor	Google Docs		https://docs.google.com/document/d/1FcO_iyvdKoVISlvzO1SUUvZivcmrUh6AVXC012Cpixo/edit SKA-TEL-SKO-0001072-02_SKAOtoSRCTechnicalInterfaceDescription
Block diagrams	draw.io		www.draw.io
Add-ons	Table of Contents Cross Reference		Used for heading numbering. Used for figure & table numbering and references.

ORGANISATION DETAILS

Name	SKA Organisation
Registered Address	Jodrell Bank Observatory Lower Withington Macclesfield Cheshire SK11 9DL United Kingdom Registered in England & Wales Company Number: 07881918
Fax.	+44 (0)161 306 9600
Website	www.skatelescope.org

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAA	Authentication, Authorisation and Auditing/Accounting
ADP	Advanced Data Product
AENEAS	Advanced European Network of E-infrastructures for Astronomy with the SKA
API	Application Programming Interface
AusSRC	Australian SKA Regional Centre
DFAP	Data Flow Advisory Panel
DID	Decentralized Identifier
IVOA	International Virtual Observatory Alliance
NREN	National Research and Educational Network
ODP	Observatory Data Product
OSO	Observatory Science Operations
RCA	Regional Centre Access
SDP	Science Data Processor (design element)
SDP	Science Data Product (an ODP or ADP)
SDPC	Science Data Product Catalogue
SKA	Square Kilometre Array
SPC	Science Processing Centre
SRC	SKA Regional Centre
SRCCG	SRC Coordination Group
SRCnet	Network of SKA Regional Centres
SRCSC	SRC Steering Committee
TAP	Table Access Protocol
TMC	Telescope Monitoring and Control (system)
UID	Unique Identifier
W3C	World Wide Web Consortium
WAN	Wide Area Network

1 Introduction

1.1 Purpose of the document

This document is a technical deliverable describing the external software interfaces of the Square Kilometre Array (SKA) Observatory with the SKA Regional Centres (SRCs).

This description covers the interfaces providing synchronisation of the Science Data Product Catalogue, bulk transfer of data products from the SKA Observatory to the SRCs, and interoperation between Authentication, Authorisation and Auditing (AAA) systems. It also identifies operational interfaces which were out of scope during the pre-construction design leading up to the Critical Design Review, and therefore will require further design effort and policy development prior to construction.

The SRC requirements [RD1] and software architecture are under development, so the software interfaces described here will evolve as the design is taken forward. For a discussion of the network infrastructure refer to [RD2].

1.2 Roles and Responsibilities

Roles and responsibilities are assigned for the purpose of progressing the interface through the SKA Critical Design Review stage and to establish a dialogue between the leading and following party. These assignments may change in future revisions.

The SKA Organisation is the Owner of the interface.

The Feature Owner and Product Owner, who were identified as part of the SAFe planning process[1], together with the co-authors, are the "leading party" of this interface.

The SKA Regional Centre Steering Committee (SRCSC) [RD4] is the following party of this interface. Contact point is the SRCSC Chairperson.

2 References

2.1 Applicable documents

The following documents are applicable to the extent stated herein. In the event of conflict between the contents of the applicable documents and this document, **the applicable documents** shall take precedence.

[AD1] SDP Architecture, SKA-TEL-SDP-00000013, rev 07
[AD2] SKA1 Interface Management Plan, SKA-TEL-SKO-0000025, rev 02
[AD3] SKA1 AAA Interface Control Document, SKA-TEL-TM-0000201, rev 02

[1] www.scaledagileframework.com

2.2 Reference documents

The following documents are referenced in this document. In the event of conflict between the contents of the referenced documents and this document, **this document** shall take precedence.

[RD1] SKA Regional Centre Requirements, SKA-TEL-SKO-0000735, rev 03
[RD2] Global and National Networks for SKA Science, SKA-TEL-SKO-0000937, rev 02
[RD3] SKA1 AAA Design Report, SKA-TEL-TM-0000037, rev 01
[RD4] SKA Regional Centre Steering Committee (SRCSC), Terms of Reference
[RD5] SKA Regional Centres: Background and Framework, SKA-TEL-SKO-0000706
[RD6] AENEAS, http://www.aeneas2020.eu
[RD7] Australian SKA Regional Centre, White Paper, v1.0
[RD8] A Primer for Decentralized Identifiers, https://w3c-ccg.github.io/did-primer/
[RD9] Registry Interfaces, http://www.ivoa.net/documents/RegistryInterface/
[RD10] NASA Systems Engineering Handbook - NASA Sp-2016-6105, rev 2
[RD11] Rucio - Scientific data management, arXiv:1902.09857v2

3 Context and Rationale

The SKA regional centres (SRC) will be regionally funded entities offering SKA data access and processing resources for use by SKA scientists, and general astronomy users [RD5]. These SRCs will collaborate to form a global network that will provide the SKA community with: access to support using the SKA and its science data products; a platform for collaborative science; a transparent and location agnostic interface for users; access to project data for all SKA users and a place for development of software tools for analysis, modelling, visualisation.

The Science Data Processors can only be accessed by Observatory operations staff. Therefore, SRCs are the only locations where users can get actual access to SKA data products. The SKA Observatory will deliver science data products from the Science Processing Centres (SPCs) into SKA Regional Centres. However, in addition to this primary interface (i.e. the bulk data delivery) there are further interactions between SKA and the SRCs which we describe here. These are still part of a developing landscape and as such cannot be treated as fixed, but we introduce them here to provide as full as description as possible, and so that gaps can be identified and filled in in future work leading to a full ICD.

The division of roles and responsibilities between SKAO and the SRCs remains under a degree of development, but can be broadly understood in Figure 1. After initial interactions between the Observatory help desk and prospective users to develop successful proposals into configured projects, the SKA Observatory will execute the scheduling blocks on behalf of a user. During this “Project Execution” phase Observatory Data Products will be generated in the SCPs. Following this, these data products will have their metadata included in the Science Data Product Catalogue and they will be delivered to SRCs.

Once products are successfully part of the science archive, and visible to appropriate users in SRCs, the project enters a “Science Extraction” stage where user-led analysis dominates the interaction with data products. The analyses in the SRCs may give rise to further data products (“Advanced Data Products”) which should be included in the Science Data Product Catalogue and accessible to users across the SRC network.

After the proprietary period has elapsed, all users will be able to use the SRCs to perform data discovery and a (probably limited) range of data analysis and extraction options.

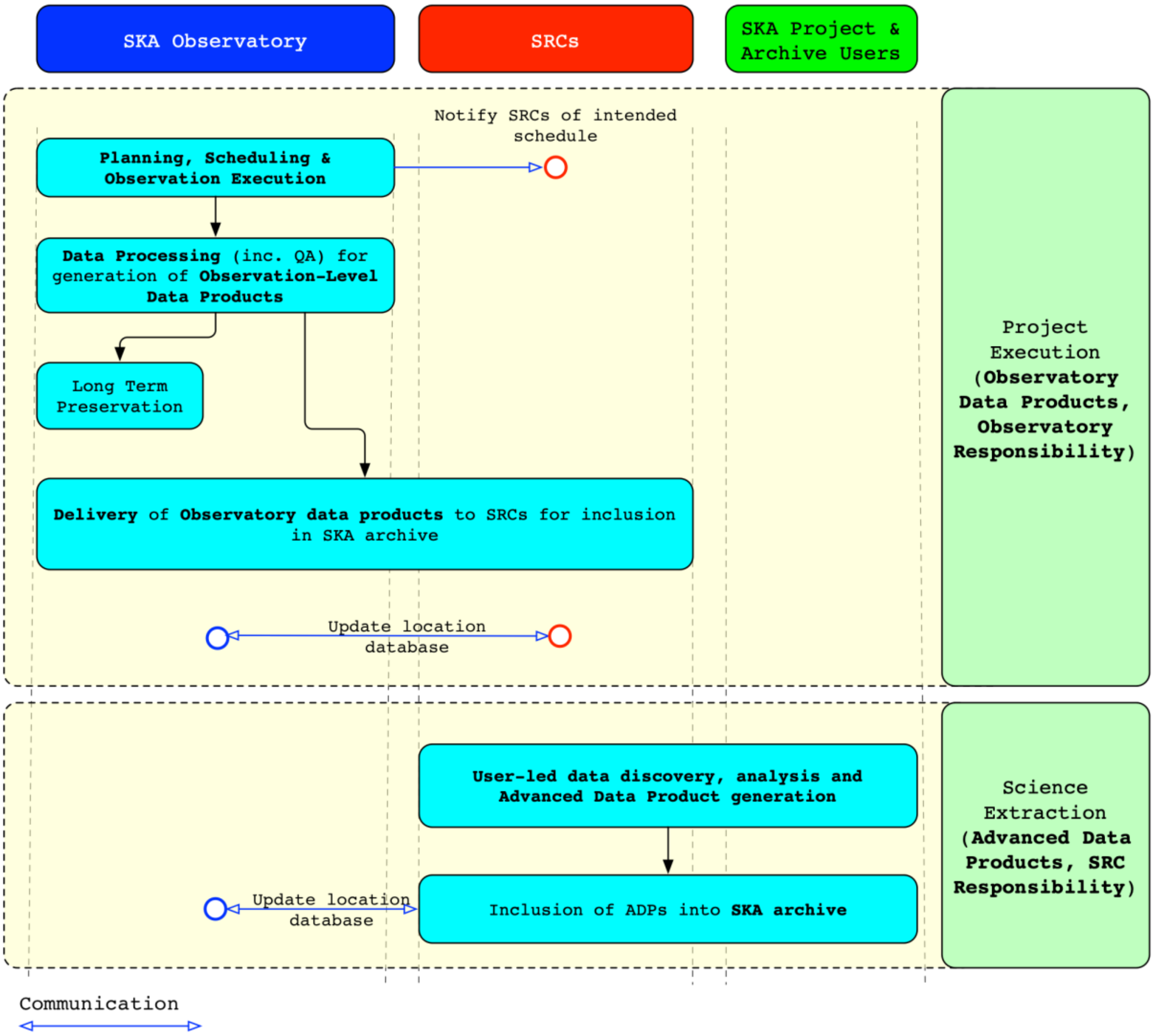

Figure 1: Roles and Responsibilities

3.1 Interface Evolution to Date

At the time of SKA1 baselining and the Preliminary Design Review a number of L1 archive requirements had been allocated to the SDP element. However, it later became clear that SDP would not include a science archive interface offering data discovery and retrieval services to the community. At that stage all L1 archive requirements were dropped and removed from SDP scope. The plan to transform archive requirements to operational requirements was subsequently revised and the Data Flow Advisory Panel (DFAP) was formed. DFAP delivered a report to the SKA Board in which it proposed a collaborative network of SKA Regional Centres (SRCs). The DFAP ceased to exist after delivering its report. Later the SKA Regional Centres Coordination Group and, following that, SRC Steering Committee [RD4] were established to progress the matter. Members of the SRCSC are appointed by SKA partners and typically sponsored by regionally-funded SRC initiatives such as AENEAS [RD6] and AusSRC [RD7]. At the SDP Critical Design Review the interface to SRCs was out of scope though.

Consequently, the SDP CDR review panel pointed out this design gap and recommended corrective action during the SKA Bridging phase, which has led to this document.

With preconstruction consortia no longer in existence the respective architecture [AD1] and in particular its interfaces are, at the time of writing, in a de facto frozen state leading up to System CDR. This limits the options to actively steer the design. As a result, this document is an interface description rather than an Interface Control Document (ICD), albeit in the spirit of the SKA Interface Management Plan [AD2] and with the SRCSC as the *following party*.

According to the NASA Systems Engineering Handbook [RD10] an interface definition document "*is a unilateral document controlled by the end-item provider, and it basically provides the details of the interface for a design solution that is already established*".

3.2 Interface Summary

Figure 2 gives an overview of the interfaces between the SKA Observatory and the SRCs. The interfaces are listed in Table 1, along with the components of the SKA Observatory and the SRCs that they connect. The SDP components are described in [AD1]. The names of the SRC components are suggestions only, since the software architecture of the SRCs is yet to be defined.

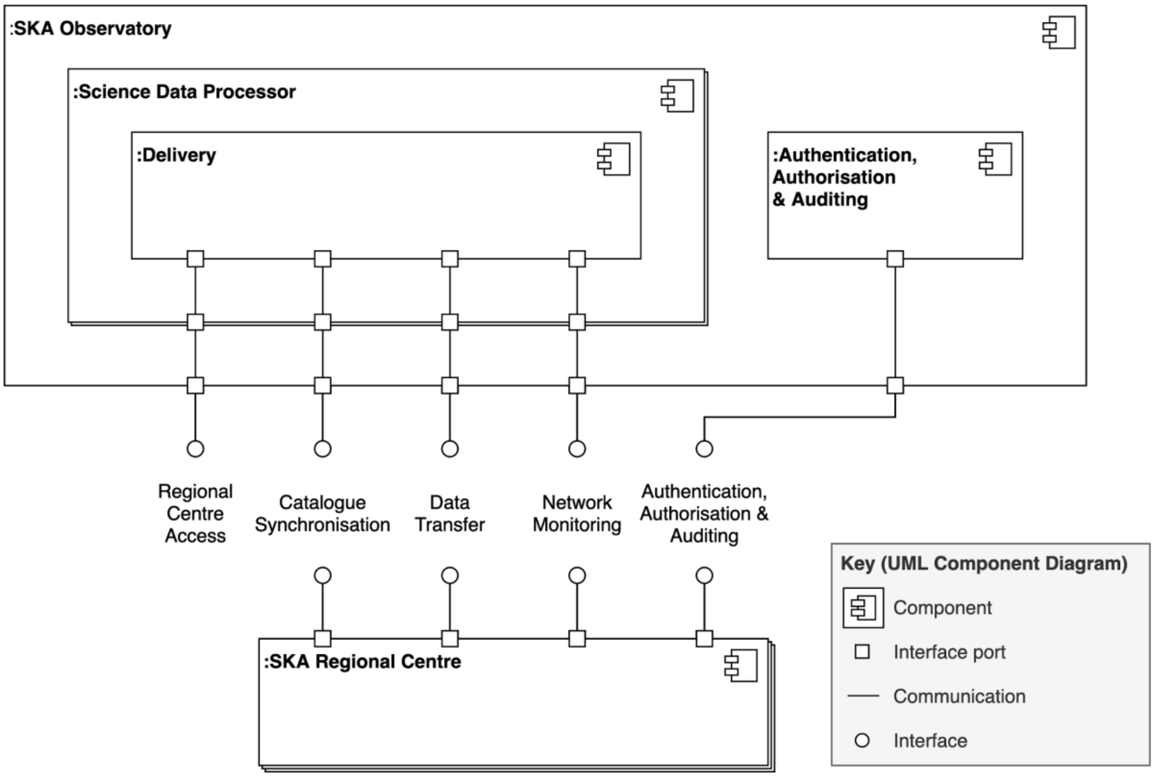

Figure 2: Context of SKA Observatory and SRC interfaces.

Table 1: Components of the SKA Observatory and the SRCs connected by the interfaces

Interface	SKA Observatory component	SKA Regional Centre component
Regional Centre Access	SDP Regional Centre Access	N/A
Catalogue Synchronisation	SDP Regional Centre Access	Science Data Product Catalogue
Data Transfer	SDP Transfer Endpoint	Transfer Endpoint
Network Monitoring	SDP WAN Health Monitor	WAN Health Monitor
AAA Interface	Authentication, Authorization and Auditing	Authentication, Authorization and Auditing

There are four interfaces between the SPCs and the SRCs shown in Figure 2. They are

- Regional Centre Access: this is an administrative interface for SRC operators to monitor the status of the SDP Delivery component and provide detailed destination information for science data products. It is not intended to interface to SRC systems.
- Catalogue Synchronisation: this synchronises information between instances of the Science Data Product Catalogue (SDPC) at the Science Processing Centres (SPCs) and the SRCs.
- Data Transfer: this is the interface by which the data products are transferred between the SPCs and the SRCs.
- Network Monitoring: this is used to monitor the state of the network between the sites.

These interfaces are described in detail in Section 4.

There is an additional interface between the SKAO Observatory Authentication, Authorisation and Auditing (AAA) component [RD3] and the equivalent component in the SRCs. This is described in Section 5.

4 SDP – SRC Interfaces

4.1 Regional Centre Access

This provides a means for operations staff at the Regional Centres to access the Delivery components at the SPCs. This enables operations staff access to monitoring information relating to the transfers that have been scheduled and subscriptions for data products that will be transferred in the future. It also enables them to set the storage targets that particular transfers should be made allowing them to control where the data is placed within their centre. The interface provides a means for operators to add subscriptions for data products that have not been setup by the SKA operators, though the ability for SRC operators to add these is based on policy set by the observatory.

4.1.1 Subscription Interface

The subscription interface is used to insert rules into the subscription database. These are simple rules that provide meta-data that is matched in the SDPC and links these to destination SRCs where the described data should be transferred to when it is available. This interface will primarily be used by SKA operators that set the overall data distribution rules, but could be made available to SRC operators for small requests to replace lost data.

4.1.2 SRC Data Landing Interface

The SRC data landing interface is used by SRC operators to set where (i.e into which of their particular storage systems) subscribed data should be transferred to at the site they represent.

4.2 Catalogue Synchronisation

ODPs and ADPs are generated at geographically distributed sites. The Science Data Product Catalogue is federated by SDP to form a global index which in turn is replicated to SRCs. Part of the catalogue are access references to ODPs and ADPs. The lookup service for the storage locations of ODPs and ADPs is called Location Service.

4.2.1 Science Data Product Catalogue

The SDPC is an index of ODPs and ADPs. An individual SRC may host a subset or all of the ODPs and may add and host ADPs. The SDPC is federated and globally-accessible with metadata about all available ODPs and ADPs. SDP adds an entry to the SDPC when releasing a science data product. The provenance of ADPs are harvested from SRCs using a yet to be defined federation protocol (TBD06).

The SDPC shall include sufficient metadata and linkage to:

- uniquely identify each ODP and ADP,
- identify the publisher of each data product in this federated multi-telescope catalogue,
- crosslink with the Location Service (see Section 4.2.2) and associate all artefacts to each data product,
- provide data provenance in compliance with Observatory Data and Data Stewardship Policies (TBD01) including those of the IVOA,

- provide software provenance to enable post-processing including linkage to snapshots of the Science Data Model [AD1] and other auxiliary data products and metadata services,
- enable access control in compliance with the Data Access Policy (TBD02) and the TMC architecture [AD3].

4.2.2 Location Service

The location service keeps track of the location(s) of each science data product including three items:
1. data product UID
2. data location UID(s) for a given data product
3. availability status for a given data product UID

The data product UID can act as a foreign key into the SDPC. Replicas share the same data product UID and have different data location UIDs. If several copies are stored in a single data location then it is up to the individual SRC to determine which copy gets accessed and how to keep copies consistent. A given SRC can have several data location UIDs. For clarity, the Location Service does not and should not require physical locations of individual storage elements. A science data product becomes *available* once it has been confirmed to be successfully delivered into the SRC network and it is included as an entry in the SDPC.

A Unique IDentifier (UID) scheme (TBD04) for Science Data Products and Data Locations applies across the network of SRCs. The responsibility for creating conforming UIDs shall lie with each SRC releasing ADPs for inclusion in the SDPC.

A candidate scheme for decentralized, self-administered identifiers is that of the DID by W3C [RD8]. Irrespective of picking a particular scheme, it is of interest to note that many (database) systems provide functionality to generate IDs which, when combined with a namespace or method, form a globally unique ID. A UID prefixed with the name of the scheme makes it self-contained and fit for long-term preservation.

For the benefit of readers concerned with the implications on SRCs we refer to two existing implementations of location services: (1) the IVOA Registry [RD9] and (2) Rucio [RD11]. The IVOA Registry is actually a standard with several interoperable implementations in the Astronomy domain. Rucio is an open source scientific data management framework that has its roots at CERN and the ATLAS experiment.

4.2.3 Failure Modes

The SDPC interface shall be able to detect and report the following failures:
- service failure
- storage failure
- network failure
- transfer/replication failure

Depending on the nature and severity of a failure the Operations Plan (TBD03) determines the appropriate cause of action including whom to notify and the appropriate means of communication.

Failure detection may either happen through self-diagnosis or platform health checking. The transfer/replication failure detection method shall distinguish between schema version conflicts and data corruption. Whether transfer or replication failure is the more appropriate term will depend on the chosen federation protocol (TBD06) and its implementation.

In this context, the term *schema* refers to the contract between transfer/replication endpoints. It is either an explicit schema, e.g. that of a database, or an implicit one like, for instance, an API implementing the replication interface. Whatever the means, it shall support the detection of schema version conflicts between replication endpoints.

4.3 Data Transfer

4.3.1 Bulk Data Transfer Interface

Since the SPCs will distribute the data products by pushing them to the SRCs, they need to know to which SRCs to send the data. This is likely to be determined on the basis of the observing project instead of individual data products. The SKA Observatory will set the data transfer policy based on the requirements of the project in question.

SRCs provide information about where to send the data products through the SRC Data Landing Interface (section 4.1.2). In addition, the Regional Centre Access interface (section 4.1) allows a response to changing operating conditions of the network or the sites receiving the data.

The bulk data transfer interface is used for sending and receiving data products. It is critical that sites run a transport protocol that is designed and optimized for large data transfers over WAN. GridFTP has been the main protocol for many years, but is being replaced by protocols such as those implemented in xrootd at many sites. The SDP data delivery sites will be using multiple scheduling systems to set up data transfer endpoints for each transfer, and for scheduling the use of the WAN link leading out to the SRCs. The scheduling to manage the setting up the transfer endpoints, though interaction with the Buffer, in which data is placed before it is moved to long term storage, is driven by the Delivery System transfer Subscription service. Once endpoints are set up, these are passed to a WAN transfer scheduler that determines how to fill the WAN pipe most effectively. FTS is a suitable tool for performing the WAN scheduling, though Rucio [RD11] could also fit this purpose. If the same interface is to be used to transfer data between SRCs they will also need to deploy scheduling tools to manage the initiation of data transfers in order to use their WAN links efficiently. SRCs must support the WAN transport protocol (TBD07) deployed at the SPCs.

The SRC operators will use the Regional Centre Access (RCA) component to setup the delivery locations for all of the subscriptions for data delivery to their site. They will be able to track the delivery schedule through the RCA and provision receiving storage appropriately. The SDP site will need to stop delivering to a particular storage location if delivery failures occur. The SRC will be informed of this though the RCA (see Section 4.3.2) and should be informed by out-of-band alerting systems (see Section 6.2).

4.3.2 Error Handling

The SDP will be able to detect when various failures happen in the delivery process via logging from the Transfer Service. These failures include networking failures, failures of systems at the receiving site, and storage space exhaustion at the receiving site. WAN failures should be identifiable using that WAN monitoring component of this interface (see Section 4.4). All failures detected will be communicated to the SRC operators through RCA and some should also be alerted through a separate out-of-band reporting system (see Section 6.2).

Table 2: Data transfer failure modes and actions

Failure	Detection	Action
Transfer node failure	Data Centre monitoring	Inform SRC operators and WAN providers through RCA and out-of-band alerting indicating if the site is still able to operate in a degraded state
WAN failure from SPC	Failure reports from Transfer Service and alert from Network Monitoring Service	Inform SRC and operators of the global engineered SKA WAN via out-of-band alerting
Connection failure to specific site	Failure reports from Transfer Service	Inform SKA WAN operators and SRC staff at impacted sites using out-of-band reporting
Write failure at destination SRC	Error reporting from Transfer Service	Inform operator at destination SRC using RCA and out-of-band reporting. Also check if SRC has an optional storage destination configured and if so start transferring to that

4.4 Network Monitoring

This interface is used by the WAN Health Monitor to test the network by contacting monitoring services located at remote sites and having the services at other sites contact them. PerfSONAR is a current widely-deployed example of such a monitoring tool. It comes in two parts, one to measure bandwidth and one to measure latency. Keeping an ongoing record of the network health using this type of active monitoring greatly reduces the time it takes to detect problems and identify their causes. Having monitoring points as close to the actual physical network interface providing the data as possible helps to reduce the time taken to isolate problems and therefore, to correct them.

This interface should be available to the NRENs providing the WAN connections used for SKA data delivery [RD2]. They will then be able to combine information collected from this interface at each SDP and SRC with the information collected on their network segments, to provide them with the most up to date information on the health of all networks that could be impacting SKA users.

Access to this information should be available to SRCs through the SRC interface. It is strongly recommended that the SKA Observatory and the SRCs work with the NREN providers to ensure that all information relevant to the WAN network used for SKA data transmission is available to their operations staff as this could greatly reduce the time taken to understand network failures and performance problems.

PerfSONAR is the suggested toolset to use on this interface. The SKAO should ensure that is still the toolset preferred by the WAN providers before finalising the SRC requirements.

5 AAA – SRC Interface

The Data Access Policy (TBD02) is enforced through the SKA Authentication, Authorisation and Auditing (AAA) system [RD3]. The AAA system supports the concept of user roles and group membership. Access control information is included in the SDPC and pertains to data and metadata of each data product. Access permissions can be granted based on user role (e.g. Principal Investigator) and group membership (e.g. by Observing Project). The AAA interfaces between Observatory systems and SRCs are described in [AD3].

6 Missing SKA – SRC Interfaces

This section identifies required interfaces between the SKA Observatory and the SRCs that are not described in the Applicable Documents and, to the best of knowledge of the authors, do not exist elsewhere in the SKA CDR design.

6.1 OSO – SRC Interface

The purpose of this interface between Observatory Science Operations (OSO) and the SRCs is to provide the SRCs with information about the observing schedule to allow for provisioning of computing and storage resources. The information is provided at two levels

- The proposed science schedule: following the process of reviewing observing proposals, SRCs are provided with information about expected resource requirements based on the accepted proposals.
- A list of executable scheduling blocks: closer to the time of observation a list of executable scheduling blocks is generated. It allows the ODP delivery timeline and SRC resource estimates to be refined.

The actual observation schedule may change for operational reasons. The interface communicates an expected or earliest arrival time of ODPs as well as upper limits for resource requirements for receiving and possibly post-processing.

6.2 Out-of-Band Alerting Service

In the SDP architecture [AD1] all alerts of problems are communicated via the monitoring and control interfaces provided by the Telescope Monitoring and Control element. Here we highlight the need for SRCs to receive alerts out-of-band in some cases. That means that the alerts are not sent over the networks used for the production work since these could have failed. Also, if there is a security breach that could impact SRCs connected to the production WAN, using this WAN to inform SRCs would also inform the attacker that they had been detected. The sites must adhere to a SKA wide policy on reacting to a security breach, which might involve isolating the problem and detecting the source of the attack before the attacker becomes aware of the detection.

The SKAO should define a set of alerting mechanisms that SRCs must adopt, including SMS to operator cell phones. This service should also have configuration options available to the SRC operators to allow multiple forms of alerting. The configurable alerting enables SRC operators to receive messages out-of-band over telephone connections, or to their own alerting services to be sure that the information is available to them in the way that suits their local working practice most effectively.

7 Other Interfaces and Services

This section identifies interfaces and services beyond the scope of the SKA – SRC interface but which are related to it and may have an impact on its design and implementation.

7.1 Interfaces between SRCs

Interfaces between the SRCs are not described in this document, but they must be defined and be compatible with the interfaces described here and with SKA policies. It is understood that SRCs may evolve from existing facilities, therefore, it is assumed that an SRC Operations Plan (TBD06) or an agreement of equivalent suitable nature will make provisions for the technical verification of service compliance. The term service shall be defined by choosing a suitable Service Model (TBD05). See also section 7.2.

7.2 Service Model

The SKAO and the network of SRCs can be modelled as a service infrastructure. This document focuses on the interfaces between services. A Service Model (TBD05) facilitates the definition of service level agreements not only with commercial partners.

The general service levels for each service and the organisational responsibilities of SKAO and SRC can be captured, for instance, using the respective templates provided by the Information Technology Infrastructure Library (ITIL). Service descriptions include interface standards, availability target, capacity, access management, responsibilities, type of support, procedures for handling exceptions, to name a few. It is important that these can be actually measured and turned into system requirements such that the quality of the service can be monitored. With that KPIs can be categorized into operational, managerial, etc.

8 TBDs and Future Work

TBDs that cannot be unilaterally resolved by the leading party of this interface and by the time of System CDR will become subject of an implementation roadmap (TBD08) involving the *leading* as well as *following party*. In this way a future revision can be elevated to an Interface Control Document.

Table 3: List of TBDs

ID	Description	Responsible at SKAO[2]
TBD01	Observatory Data and Data Stewardship Policies	SKAO Operations / SKAO SRC Scientist
TBD02	Data Access Policy	SKAO Operations
TBD03	SRC Operations Plan (or equivalent contract/agreement)	
TBD04	Unique IDentifier (UID) Scheme	SKAO SW Architecture Team
TBD05	Service Model and KPIs	TBD08
TBD06	SDPC Federation Protocol	TBD08
TBD07	WAN transport protocol	TBD08
TBD08	SRC Implementation Roadmap	SKAO Operations

[2] The SRC involvement in resolving the tabled items is a TBD in itself.

Appendix

Recommendations received from the System-CDR panel as well as feedback from the SRCSC are replicated in this appendix, together with responses by the authors. The actual advancement of this interface description to an interface control document is planned after CDR close-out in a future revision.

A.1 S-CDR Panel Feedback

OAR SCO-49 SKAO - SRC Interface Description is/should be a formal ICD

OAR Description text taken from Jira ticket:

> "However, in addition to this primary interface (i.e. the bulk data delivery) there are further interactions between SKA and the SRCs which we describe here. These are still part of a developing landscape and as such cannot be treated as fixed, but we introduce them here to provide as full as description as possible, and so that gaps can be identified and filled in in future work leading to a full ICD."
>
> It is not clear to me if this Interface Description Document will form the basis of a bona fide ICD (which is stated as intent in this document, see above). Since this is out of scope for the System CDR as I understand it, this ICD should be reviewed at a "post-CDR" review of this interface at [sic] soon as possible in the future.
> The SKAO - SRC interface is a critical one for SDP, and has the potential to generate requirements creep if this interface is not well-defined early on. While it is not clear how many and which institutions will host SRCs, it is nevertheless important to constrain this interface in exactly the same way as other interfaces. In order to do this, the SKA Regional Centres Coordination Group and SRC Steering Committee should be maintained to speak on behalf of potential SRC hosts and ensure open issues are fully addressed in this ICD.
> In addition to sections 6 and 7 of this document, which identifies open issues, here is a summary of interface requirements gleaned from this and other baseline documents with unstated or potential requirements creep areas:

#	CDR OAR Panel Feedback	Response
1	How does SKA handle it if the SPC - SRC network is down for a period of time? Will there be "catch-up" capacity or prioritization scheme, for example? How many SRCs can be simultaneously fed the data? Is the network monitoring for the observatory sufficient, too much, or too little for the SPC - SRC network?	SPCs have a dump interface that can be used to transfer data to portable storage systems than can then be shipped.to appropriate SRCs. Which SRC would receive a particular delivery would depend on what part of the overall network has failed. Each SPC can send data at the rate enabled by their network interfaces. The important question is how much data can be moved, not how many SRCs can be served simultaneously. The data distribution policy will determine which SRC receives each product from the SKA. SPCs are specified with a 100Gb/s interface

		at this time. If that proves not to be sufficient due to changes in the way that the telescope operates over time, adding additional 100Gb/s interfaces should be possible, though that would depend on being able to fund the WAN links required to support additional bandwidth. Monitoring will mainly be performed by the network operators providing the different network segments. SPCs should ensure they have monitoring at the edge that will interoperate with the monitoring provided by the network operators. This will enable a view of all segments to the SKA operators and to all regional network providers enabling problems to be identified quickly. Note that this is how the CERN networks operate so most of the network operators that SKA will work with already use this approach.
2	Will SRCs be satisfied without some form of on-demand or interactive query of the archive? Is the plan for catalog synch frequent enough?	The SRC operators will all be able to query the catalogue to see what data will be available. Most queries should be performed on replicated versions on the catalogue, but it is possible for the SRC operators to query the primary copies through the SRC interface at the SPCs. It would be a matter of policy if they would be given access to this information, not an architectural issue. The rate at which the catalogue is synchronized would be adaptable and would depend on network availability to some extent, though data flow rates can also be adjusted to ensure replication rates that work for the SRCs. As pointed out there is SRC operators could be provided with access to the primary copies if replication rates are not satisfactory for the most recent observations. The SKAO has the capability to generate VOEvents for time critical notifications.
3	What mechanisms and to what degree will it be possible to federate data across the SKAO and SRCs?	The location of the initial copy of a set of ODPs will be defined by the SKAO, possibly as part of the proposal preparation process. The rules for placement of additional copies will need to be agreed to by the SRC Entity[3]. In all cases the location of data will be available through the location service (4.2.2). The ability to obtain copies of the data from other SRC sites will need to be covered in an MoU. This MoU could also include the ability to execute code at an SRC site holding the data. The availability of

[3] The SRC Entity is defined in a White Paper by the SRCSC which is in draft state as of Feb. 2020

		software containers has made the process of moving code to data much easier, so the main issues will be having suitable agreements in place to enable this to happen, along with the access enforcement implementation which all SRCs must employ. The other part of federation is the authentication and authorisation services that provide the information needed by the access enforcement implementation.. These are now well catered for using either a Federated Identity Management or using grid security tools. Also, section 4.2.1 has a bullet list of items that enable federation, including data and software provenance. As such the federation capability is proportional to the investment in implementing these items. The IVOA registry standard[4] is a possible, well established federation mechanism. It allows for seamless integration of SKA data services and data collections with existing archive portals and tools[5].
4	Will the SRCs play any role in disaster recovery for the SPCs or other SRCs?	This is a matter of policy. The architecture allows for this.
5	The variant of the SDP software that is to be provided to the SRCs may entail significant porting work if the SRC infrastructure is allowed to vary widely from the SPCs, which may occur if the SRC infrastructure is constrained by SRC host institution standards independent of SKA. Who is responsible to ensure the SDP variant is compatible at the SRC? How can this be constrained to fit within allocated resources to prepare these deliverables?	We believe the SDP software will only be supported for a small number of the most relevant architectures. The SKA will publish this information to enable sites to make purchases that will be able to run the SDP software. It should be noted that CERN had very tight restrictions on the hardware that was acceptable, but with advances in virtual machine and software container technologies, these restrictions don't need to be as tight as they were when CERN was starting up. Clearly the SKAO can't be responsible for supporting all hardware solutions and accelerator technologies that may become available.
6	It is likely that Science Data Product archive at the SRCs will be derived from SDP ODP archive, but	There is a need for requirements gathering of what is required as far as extensions to archive software required at the SRC sites, though this can gain input

[4] http://www.ivoa.net/documents/RegistryInterface/
[5] http://www.ivoa.net/astronomers/applications.html

	considerably adapted for user access. This will potentially require non-trivial interaction between SDP archive designers and SRC archive designers. How can this be resourced and constrained to fit within allocated resources?	from the various regional projects that have been active over the past few years. Often funding can be sought from regional funding bodies to support this type of development work, though coordination between SRCs will be required.
7	Cyber security and AAA at an SRC will likely have to fit within an overall institutional cyber security program and environment. What is the architecture of AAA and how hard will this be? Who is responsible for what parts of this integration?	This is being handled by the SKA. The architecture is based on the widely accepted Federated Identity Management approach.
8	The SRCs are designated to play a role in acceptance of SDP data products. (SKA-TEL-SKO-0001075-01_DesignBaselineDocument-part2forsignaturesigned.pdf p564). What is this role and what implications does this have on technical capabilities?	The choice of words in the referenced text on p. 564 might be misleading: "Accepting" is about receiving and ingesting the data products.

A.2 SRCSC Feedback

Preamble by SRCSC commenter:

In general I think this is a good starting point for a discussion on interfaces. To me it highlights a few high level and basic issues we need to develop further:

#	SRCSC Feedback	Response
9	The SKAO-SRC interface is an interface between two operational systems. It is not an interface between a user and a service provider only. Both sides of the interface will have operational processes that	The term SRCnet or SRC network could be used in place of SRCs, wherever it is accurate to use the singular form.

	connect at various times and with potentially different directionality and latency.	
10	The document assumes that the SRC interface is relevant to an individual SRC and hence that the processes involved are from SKAO to multiple but individual SRCs each with a common interface. As noted in point (1) and in our discussions on the need for an SRC operational entity, I think we need to view the interface as a single interface between SKAO operations and the SRC operational entity. This will simplify the SKAO operations process and allow SRC operations to manage the SRC-facing "data traffic" and "operational process flow" in an optimal manner. This will mean SKAO can execute operational processes when required/ready without the need to monitor the operational status of individual SRCs and the SRC operational entity can buffer/manage as required across a collection of SRCs with differing capabilities and availability at any given time.	Section 7.2 mentions the possibility to adopt a service model to aid with the definition of homogeneous interfaces and responsibilities across SKAO and SRCnet. This appears in line with the intent of the comment. As far as technical interfaces are concerned the cardinality depends on the architecture - like peer-to-peer, client-server and micro-services to name a few - and the abstraction level. Taking responsibility for data resilience, for instance, can be a matter of defining a backup and recovery mechanism once (availability, capacity, procedures etc.). It may require several instances supporting it at various data locations (4.2.2) and each data location may start and drop support individually. This goes to show that a 1:1 as well as 1:n relationship may both apply in a given context. When evolving the document to a full ICD the cardinality needs to be clearly expressed.
11	There is more than one SKAO-SRC interface. The one under discussion is the "technical" one associated with data flow/management. The other one is a more "human" interface associated with TAC process and resource allocation as we have discussed in our first F2F. The resource allocation interface will require some tool and database support and will, I assume, result in a technical interface as part of operational tools and processes associated with TAC, Phase 1 Proposal Preparation etc.	This is an omission. It is understood that the TAC requires information about allocatable resources. The TAC may also require a tool similar to the SDP parametric model (https://github.com/ska-telescope/sdp-par-model) to estimate the impact of altering a proposal on both sides of the interface, i.e., the SKAO as well as SRCnet.

Preamble by SRCSC commenter:

There has been an evolution driven by the Operations model and documented in The Users and their needs - white paper section 3, the result being the interaction diagram now looks quite different from Figure 1. There is now the concept of Observation-Level Data Products (OLDPs) and Project-Level Data Products (PLDPs) which are both included in the Observatory Data Products (ODPs) definition. OLDPs will be delivered from the SPCs to SRCs (I apologize for the acronym soup) whereas PLDPs will be generated in the SRCs. Because PLDPs are ODPs, they are the responsibility of SKA to be produced and QA'ed. The implications are:

#	SRCSC Feedback contd.	Response
12	PLDPs, like ADPs, will need to be added to the archive and the location service.	Naturally, the figure should be updated as the interface design evolves. It should be noted that Figure 1 in its current form allows for the inclusion of data products generated by SRCnet.
13	There will need to be an interface from SKA Observatory to SRC for SKA staff: I. to process and QA the PLDPs, implying a control and monitor type of interface; and II. to install and update processing pipelines in the SRCs; and III. to request computing and storage resources to support the processing	Given that any SRC interface can be used by SKA staff at some point, there is the potential of scope creep. Therefore, it would help to clarify whether this is merely about distinguishing a particular user role (SKA staff) or whether specific interfaces in addition to those provided to non-SKA staff are required? We have not seen requirements for software updates to be managed from the SKAO side. This could be done in a number of ways such as using CVMFS, or by local scripts pulling the versions of software required for a pipeline run from some other repository. As far as job submission, if this is going to be done centrally, a meta-scheduler could be used to farm out jobs to local schedulers at the SRCs. However, it is likely that the SRC will be processing work based on the data received, so a simple framework could be provided that checks for data available and queues jobs locally. With both items (II) and (III) some more detailed requirements on how the SRCs are expected to operate would enable the interface to be extended to support these. For (I), interfaces are provided for that in the SDP software, though the interfaces for it are outside of the SDP scope, so some investigation of how they could be provided to the SRC sites would be required.