Journal of Data, Information and Management (2021) 3:251–260
https://doi.org/10.1007/s42488-021-00055-1

ORIGINAL ARTICLE

RapidXfer - data transfer framework for square kilometre array

Priyaa Thavasimani[1]

Received: 26 January 2021 / Accepted: 7 June 2021 / Published online: 28 July 2021

Abstract
Data Management of Astronomy Data is often a laborious task and it is even more challenging for the extraordinary amounts of data expected from the world's largest radio telescope, Square Kilometre Array. There are overt issues in transferring the voluminous `SKA` data and the traditional data transfer methods are fragile especially for the data transfer between two continents. To address this, a new data transfer framework `RapidXfer` is proposed and the data transfer is achieved using two steps: international and local transfers. The efficiency of different end-to-end data transfer tools used in `RapidXfer` is evaluated on different dataset sizes. Further, a comparative study of two IRIS grid data transfer methods is made to understand each methods' advantages and disadvantages. This study can be used as a reference for the development of future SKA's data transfer operations.

Keywords Big data · MeerKAT · Square kilometre array · Telescope · Astronomy

1 Introduction

Square Kilometre Array (SKA) comes with huge challenges because of the underlying fact that the data generated will be highly voluminous. It is estimated that SKA observatory will generate 600 PB of calibrated science data products each year (Array 2019). SKA's host sites are located in South Africa and Australia, which will be integrated into phase 1 of SKA. Initial data products from these host sites are generated at different rates. Science Data Processors (SDPs) ingest huge amounts of data, which will be sent to SKA Regional Centres to provide the user community with data access for further processing. This study focuses on data transfer mechanisms and reducing the data transfer time for extremely voluminous inter-continental data transfers. The framework `RapidXfer` takes advantage of the open-source Globus transfer service and Data Management Library called Grid File Access Library (GFAL) to improve transfer speeds. The data transfer task is achieved in 2 stages i.e international and local transfers. The international transfer from the South African IDIA site to UK's Manchester HighMem (intermediate node) is achieved using Globus. The `local` transfer between Manchester HighMem (intermediate node) and DiRAC's LFC (LCG File Catalog) is achieved through Grid File Access Library (GFAL) and DiRAC's `register` command. The functionality was tested on different dataset sizes ranging from 72 GB to 1.5 TB datasets. It is evident from the results that the choice of using Globus online for international transfers from South Africa to the UK's Manchester HighMem node is ideal compared to traditional transfer using SCP. Both GFAL-Copy (with DiRAC's register) and DiRAC's add file functionality can be used for local-transfer of terra-byte scale MeerKAT datasets, pros and cons of these methods are discussed further. The paper is organised as following sections:

(1) Section 'Literature Review' discusses about SKA, SKA's data rate, SKA precursor telescopes: MeerKAT and ASKAP and their data rates, IRIS computing resources.
(2) Section 'DiRAC Data Management System and 'Add File' Functionality' includes details of the Data Management System available to manage MeerKAT data. This section also includes details of user commands used to add the file to IRIS storage.
(3) Section 'Globus and Grid File Access Library (GFAL)' explains data management tools and services of the proposed framework RapidXfer i.e. Globus and GFAL.

✉ Priyaa Thavasimani
priyaa.thavasimani@manchester.ac.uk

[1] The University of Manchester, Manchester, UK

(4) Section 'RapidXfer' - MeerKAT Data Transfer Framework' discusses the implementation of RapidXfer with Globus and GFAL.
(5) Section 'Evaluation' includes evaluation of Globus, GFAL, DiRAC Add File, and RapidXfer on varied datasets of sizes ranging from 72 GB to 1520 GB.
(6) Conclusion and Future Work

2 Literature review

The Square Kilometre Array (SKA) is a global project with requirements derived from thirteen high-priority science objectives (Dewdney et al. 2009). The SKA's telescopes will have a collecting area of one million square metres, i.e. a square-kilometre. Australian Square Kilometre Array Pathfinder(ASKAP) and South Africa's MeerKAT are the precursor telescopes of SKA that will be integrated into phase 1 of SKA. Phase I of the SKA will consist of two telescopes i.e. SKA1-MID located in South Africa's Karoo region, and SKA1-LOW in Western Australia's Murchisonshire. It is expected that the data rate between between the antennas and the correlator will be 23 Tb/s whereas the data rate between the correlator and the HPC (provided by Science Data Processors SDPs) is 14 Tb/s, which is equivalent to 12,000 PB/month (JBC for Astrophysics 2009). The final data rate will depend on the overall system to be built.

In this work, MeerKAT data is used to evaluate the performance of the proposed Data Transfer Framework 'RapidXfer'. MeerKAT is one of the precursor telescopes to SKA1-MID which will be integrated into Phase - 1 of SKA. MeerKAT consists of 64 antennas in the Northern Cape of South Africa. The expected input data rate of the MeerKAT's SDP (Science Data Processor) is about 4 terabits per second (4000 gigabits per second) (SARAO 2018). Rucio (Barisits et al. 2019), which provides a generic scalable approach to transfer data for high-energy physics experiments is still being evaluated for SKA.

UK's SKA user community process the SKA's data using STFC's IRIS Computing resources. The Science and Technology Facilities Council (STFC) is one of the UK Research and Innovation (UKRI) councils, which plays a leading role in many global science projects including the Square Kilometre Array (SKA) (2019). STFC's IRIS (2021) infrastructure is a coordinating body which provides digital research by working with infrastructure provider i.e. Worldwide LHC Computing Grid (WLCG) (2006). Grid for UK Particle Physics (GridPP) collaboration is UK's involvement with WLCG. GridPP was initially created to provide computing resources to process and manage LHC's (Large Hadron Collider) Physics data. GridPP (T.G.C. 2020 2020) plays a key role in large-scale LHC experiements by a providing number of services, which have been expanded for use by other research communities (Bauer et al. 2015). DiRAC (Distributed Infrastructure with Remote Agent Control) (T.G.C. 2020 2020) is a GridPP's framework that provides services for workflow management (WMS) and data management system (DMS) (Britton et al. 2009), where WMS facilitates running JDL (Job Description Language) jobs on the grid and the DMS provides store/retrieve access to large scale storage systems. The Data Management System(DMS) consists of 3 main components: File Transfer Service (FTS), File Catalog, and Storage Element (SE) (Tsaregorodtsev et al. 2004).

Every day, hundreds of thousands of files are transferred for scientific research purposes. Depends on the research field, the dataset size varies from few Kilobytes to Terabytes. Most MeerKAT datasets vary from a few hundreds of GB to 1 TB and the dataset is usually called as Measurement Dataset (MS) which has its hierarchy of subfolders. Often, the data generator and the processor are not located in the same place and so it is essential to transfer from the source to destination for further processing and analysis. The choice of transfer protocol is not important if the dataset size is as small as 2 GB. For the extremely voluminous datasets like SKA, the choice of transfer protocol plays a critical role. For transferring files, there are conventional SSH-based file transfer commands including `SCP`, `Rsync`, and `SFTP` which are preferable if we transfer within the same network. The Secure Copy Protocol (SCP) is used to securely transfer data from one device to another over a network. Like the FTP, the SFTP (Secure File Transfer Protocol) provides an interactive file transfer service where all the transfer operations are carried over an encrypted transport. For files with complex hierarchies, it would be appropriate to use `Rsync`. Rsync also helps to check timestamp and file size. GridFTP is an extension of File Transfer Protocol (FTP), which is widely used in grid environments.

3 DiRAC data management system and 'add file' functionality

DiRAC (Haen et al. 2015) offers data handling operations for small and large user communities. DIRAC's Data Management System can be configured to make use of FTS3, which is the file transfer service used to distribute the majority of LHC (Large Hadron Collider) data. This facilitates scheduling and monitoring efficient transfer of large amounts of data between Storage Elements. Data transfers are performed using third party services including FTS3. FTS3 uses GFAL (Grid File Access Library) as the underlying data management library. DiRAC offers a command-line facility to upload a file to a Storage Element

```
dirac-dms-add-file <LFN> <FILE> <SE>
```

Listing 1 DiRAC add file

and register it into the DIRAC File Catalog. This process can be achieved using 2 approaches either Listing 1 or 2. LFN refers Logical File Name, FILE refers local file and SE refers Storage Element.

Second way to add file to storage element is shown in script Listing 2. All the DiRAC functionality can be used through the DIRAC API. One of the DiRAC API classes is 'addFile' which is used to add a single file to Grid Storage. Usage format of 'addFile' functionality is shown in Listing 2.

Singularity containerised DiRAC Environment is set up on South Africa's IDIA to add a file directly to UK's DiRAC's file catalog by the above methods 1 or 2. It works for smaller files but breaks due to time-out and security issues for larger terra-byte scale files.

3.1 Globus and grid file access library (GFAL)

The proposed data transfer framework 'RapidXfer' combines the advantages of the open-source 'Grid-FTP' based services *Globus* transfer service (Globus 2021) and `gfal2` (Kiryanov et al. 2015), to improve transfer speeds. The advantages of using these services are explained below:

(1) Globus: Globus, a cloud-hosted and non-profit service created at the University of Chicago aid Researchers in Data Transfer through secure and reliable Software as a Service system (globus.org 2012). It can move files through HTTP or GridFTP protocol (Liu et al. 2017). Globus Transfer streamlined the process of secure transferring monumental data sets between two Globus users (Allen et al. 2012). It is credible, impregnable, highly executable, and easily usable. As per the 2014 report, more than 18000 Globus users have nearly transferred 52 Petabytes of Data (Chard 2014) and it increased to greater than 250 Petabytes and the number of users skyrocketed to 60000 in 2017 (Chard et al. 2017). To use Globus transfer, one must have source and destination Endpoints. It acts as a facilitator for data transfer allowing an endpoint to create a safe link with one another in fact the Data never travel through Globus. Researchers can make use of this service without installing any software and start the transfer process through Desktop web browsers in Firefox, Google Chrome, Safari, or Edge. The interface is user-friendly, and we can see the status of transfer with task identification i.e. `Task ID`, and once the task is completed it gives email notice. The ratio of the total number of transferred bits to the total time taken (including the retry time, downtime on the endpoints, time that the transfer is paused for credential renewal, and checksum calculations time (globus.org 2021)) is the effective transfer rate in Globus. A sample successful transfer of 1.33 TB Data is shown in 3 with an effective transfer speed of 14.78 MB/Sec.

```
>>> dirac.addFile(<LFN>, <FILE>, <SE>)
```

Listing 2 DiRAC add file

(2) Grid File Access Library - GFAL: GFAL (Kiryanov et al. 2015) is a data management library used for the transfer of data between CMS sites. It is a C library and GFAL-2 is version 2 of GFAL which simplifies the file operations in a distributed environment. GFAL-2 is a plugin-based library for file manipulation supporting multiple protocols including Webdav/https, GridFTP, xroot, SRM. In RapidXfer, GFAL2 is used and GFAL2 Util provides command line features including gfal-copy. 'gfal-copy' is used to copy a source file into a destination. Unlike DiRAC's Add File, the GFAL-Copy only copies the file from source to destination without registering the file in DiRAC's File Catalog. To use the file for processing in the IRIS DiRAC jobs, the file needs to be registered at the DiRAC's File Catalog. This process is automated in script Listing 5. The process of the 'gfal-copy' and 'register' method can be comparable to the DiRAC's Add File functionality and it is called 'GFAL-Copy+Register'. The 'register' function takes only a fraction of seconds (for both smaller and larger files) and so the time taken to achieve this is negligible.

3.1.1 Advantages of proposed 'RapidXfer' framework

Globus and Gfal serves distinct purpose as mentioned previously. Globus is designed for international transfer, but does not have the functionality to transfer the data directly to the end IRIS destination system for processing. Also, GFAL does not support international data transfers but supports transfers between only CMS sites (2018). RapidXfer automates the data transfer process by integrating Gfal and Globus. Consequently, incorporating both the tools using RapidXfer framework achieves the overall benefits of the intercontinental data transfers as well as local IRIS system transfers with higher efficiency and data rates at 40 MB/sec approximately. Thus, it increases the throughput, effective use of resources and can be replicated in any other IRIS systems (STFC-IRIS 2021). For instance, Rucio uses FTS3 (File Tranfer Service) which in turn uses GFAL as underlying file access library. FTS3 as mentioned in

(Ayllon et al. 2014), used in Rucio can transfer data with maximum throughput of 5 MB/sec. In contrast RapidXfer transfers data with a maximum throughput of 40 MB/sec which is 8 times faster than the Rucio. Presently, there is no existing framework to integrate the Globus and GFAL to get better data transfer rates. Quite a few research works in IRIS data management has been initiated especially in larger projects like SKA, yet the research works are in a nascent stage. Hence, RapidXfer framework is the forerunner in data transfer process as it integrates the available resources effectively rather than developing new tools from scratch.

3.1.2 Applications of 'RapidXfer' framework

RapidXfer is used as a study in SKA to verify the data transfer between South Africa's IDIA site to UK's IRIS grid storage. SKA host site is also located in Australia (U. SQUARE KILOMETRE ARRAY 2019). Consequently, it can be used to transfer the data between any SKA sites for instance SKA host site Australia to IRIS storage or South African SKA host site to IRIS storage and vice versa. Furthermore, many institutions across the world uses IRIS grid storage namely CTA-UK, CCFE-UK, LSST-UK, DUNE, CASU, LSC, ATLAS EXPERIMENT, CLF, CERN, Gaia-UK, ISIS Neutron and Muon Source, UK ALMA REGIONAL CENTRE, e-MERLIN, Euclid Consortium, diamond, LZ, Wide Field Astronomy Unit and LHCB-UK (STFC-IRIS 2019). These institutions generate thousands of Petabytes of data and must be transferred to distant locations for the data processing (Liu et al. 2010). RapidXfer can be easily integrated into these IRIS systems to transfer the data competently.

3.2 Contributions

(1) The main contribution is a novel framework 'RapidXfer' for transferring the terabyte-scale (> 1 TB) MeerKAT datasets from the South Africa's IDIA Ilifu cloud to UK's IRIS grid storage.
(2) 'RapidXfer' is implemented by automating the data transfer tasks using Globus dedicated endpoints for international transfers and Grid File Access Library for local transfers.
(3) The data transfer tools 'Globus' and GFAL-Copy+Register are tested and evaluated on varied data sizes ranging from small-sized datasets 72 GB to big-sized datasets 1.5 TB datasets.
(4) Efficiency of GFAL-Copy+Register and 'DiRAC's data management system feature - Add File' are also tested.
(5) The proposed framework 'RapidXfer' is also evaluated in terms of data size and effective speed.

```
$./globusconnectpersonal -setup --no-gui
```

Listing 3 Creating globus endpoint

4 'RapidXfer' - MeerKAT data transfer framework

Figure 1 visualises the end-to-end data transfer from South Africa's IDIA site to UK's IRIS grid storage. MeerKAT datasets are of different sizes mostly ranging from 1 TB to 1.5 TB. IDIA (Inter-University Institute for Data-Intensive Astronomy (2015), provides data storage and a data-intensive research cloud facility to service the MeerKAT science community in South Africa. Initially, MeerKAT data from the SKA-SA (SKA South Africa) CHPC (Centre for High Performance Computing) is transferred to the IDIA Data centres using Dedicated Data Endpoint (Aikema et al. 2020). It will be necessary to send the data to SKA Regional Centres (SRCs), so the worldwide end-user community can process the data further. SKA Regional Centres facilitates the transfer of data from SKA telescope sites to CERN's Tier 1 sites and further to other Tier 2 sites. For the UK User community to process the data on the IRIS infrastructure, it is necessary to transfer the data onto grid storage and register it within the DiRAC's Logical File catalog, where each data file can be accessed using the dedicated Logical File Name (LFN). This data transfer is achieved through 2 stages. Firstly, the data from IDIA is transferred to a local grid UI machine at the Manchester Tier2 using a dedicated Globus Endpoint 'ManchesterUI'. The source end point for IDIA i.e. Ilifu DTN and destination endpoint i.e. 'ManchesterUI' are shown in Fig. 2. The source endpoint for IDIA i.e 'Ilifu DTN' can be accessed through successful authentication.

The destination endpoint 'ManchesterUI' can be created by the command Listing 3.

The endpoints are activated by the command Listing 4.

Figure 3 shows the details of a successful Globus transfer of 1.33 TB dataset from the IDIA's Ilifu endpoint to the ManchesterUI with an effective speed of 14.78 MB/Sec. Each transfer task is assigned a `Task ID`. Figure 3 also shows when the transfer request is submitted, when the transfer is completed, how many files & directories are transferred.

After the Globus transfer from `IDIA` to the intermediate node `ManchesterUI`, the Grid File Access Library is used to transfer from ManchesterUI to a physical location,

```
$ globus endpoint activate $<endpoint id>
```

Listing 4 Starting globus connect personal

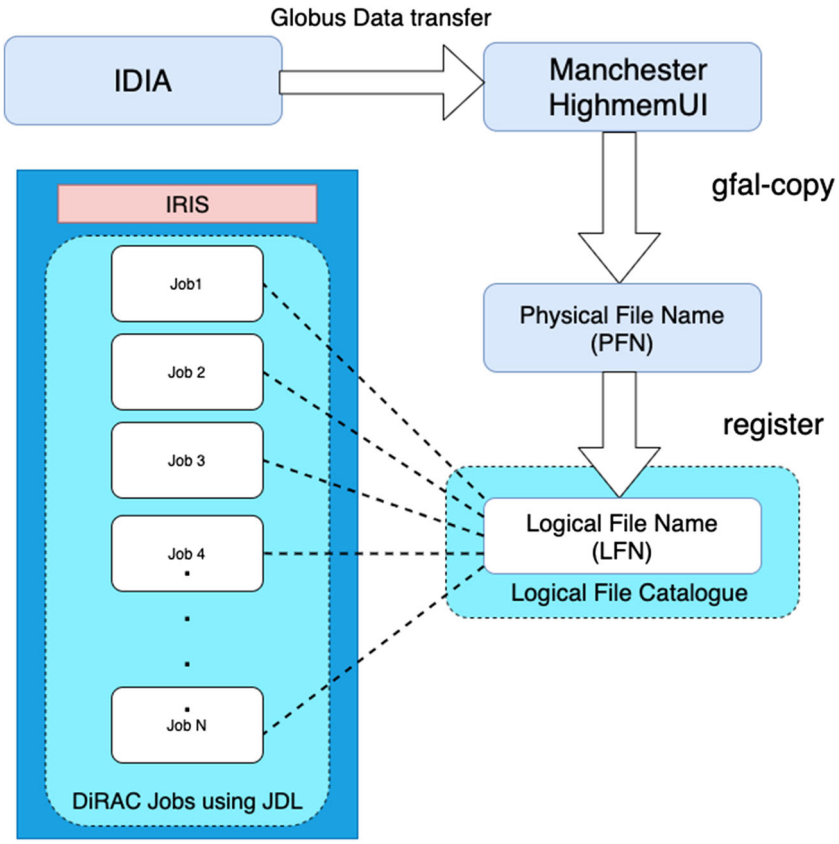

Fig. 1 RapidXfer data transfer framework and processing

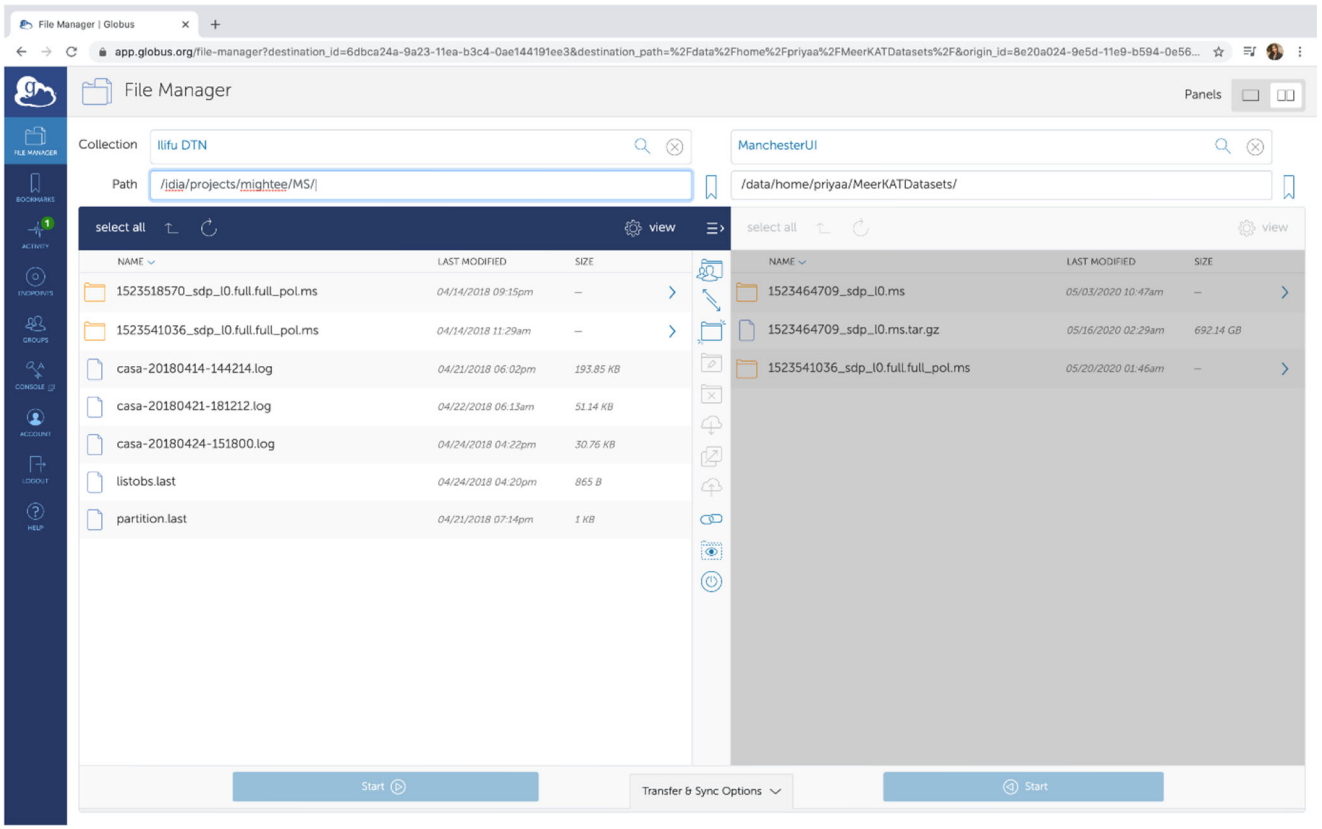

Fig. 2 Data transfer from source IDIA's Ilifu to destination ManchesterUI

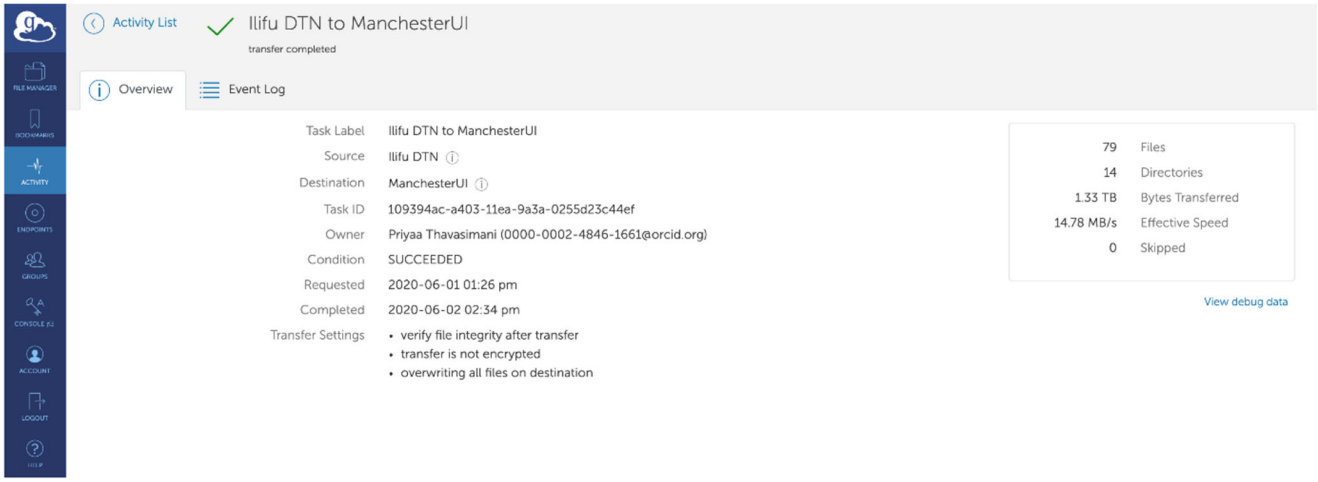

Fig. 3 1.33 TB globus transfer

then the file is given a Physical File Name, and DiRAC is used to register the file in the Logical File Catalogue. This process is automated by script 5.

5 Evaluation

Secure Copy Protocol (SCP) was used to transfer 1.33 TB of MeerKAT Data. It took 4320 minutes (3 days) from the South African site to UK's Manchester HighMem node, and the data transfer rate is 5.13 MB/Sec approximately. Further, GFAL and DiRAC's add file functionality are tested. For all the evaluations, the same MeerKAT datasets of sizes 72.82, 277, 324, 325, 336, 396.06, 471, 1300, 1330, 1520 in GB are used for consistency and, the dataset sizes ranges from 72 GB to 1.5 TB. Evaluation is done in a wide range of scenarios including testing Globus transfers, GFAL, DiRAC's add file functionality, and finally the full framework 'RapidXfer' with respect to data size and effective speed.

```
#!/bin/bash
#1 The local file that has to be uploaded
date
source <gfal-setup-file>
which  gfal-copy
voms-proxy-init --voms skatelescope.eu
gfal-copy -t 7200 file://$PWD/$1
     <PFN Location>
du -b $1 > submit.txt
value=`cat submit.txt`
filesize="$(cut -f1 <<<"$value")"
echo $filesize
dirac-dms-filecatalog-cli << eof
register file <LFN location>/$1 <PFN
Location>/$1 $filesize <Storage Element>
exit
date
```

Listing 5 GFAL-Copy+Register

(a) Efficiency of Globus with respect to SCP

Figure 4a refers to the Data transfer rate arranged from smallest dataset to largest dataset size. The smallest dataset taken for this evaluation is 72.82 GB whereas the largest dataset taken is 1520 Gb (i.e. 1.52 TB). The maximum speed achieved is 35.58 MB/Sec for the 72.82 GB dataset, and the minimum speed achieved is 14.78 MB/Sec for the 1.33 TB dataset. These speeds are comparatively far higher than SCP (Secure Copy Transfer), which took 3 days i.e. 4320 minutes to transfer the same 1.33 TB dataset and the SCP data transfer rate is 5.13 MB/Sec approximately. The Average Transfer Speed for Globus is 21 MB/sec, which is shown in Fig. 4b.

(b) Efficiency of GFAL-Copy+Register and DiRAC's Add File Functionality

Although, DiRAC's Add File internally uses GFAL library for the file transfer, due to additional overhead and checksum functionality of DiRAC's Add command 1 and DiRAC Add File API 2, the file transfer speed is comparatively lower than that of direct usage of GFAL library. The comparison between GFAL-Copy+Register and DiRAC's add file functionality is depicted in Fig. 5. The highest, lowest and average data transfer rates for the GFAL-Copy+Register and DiRAC's Add file functionality

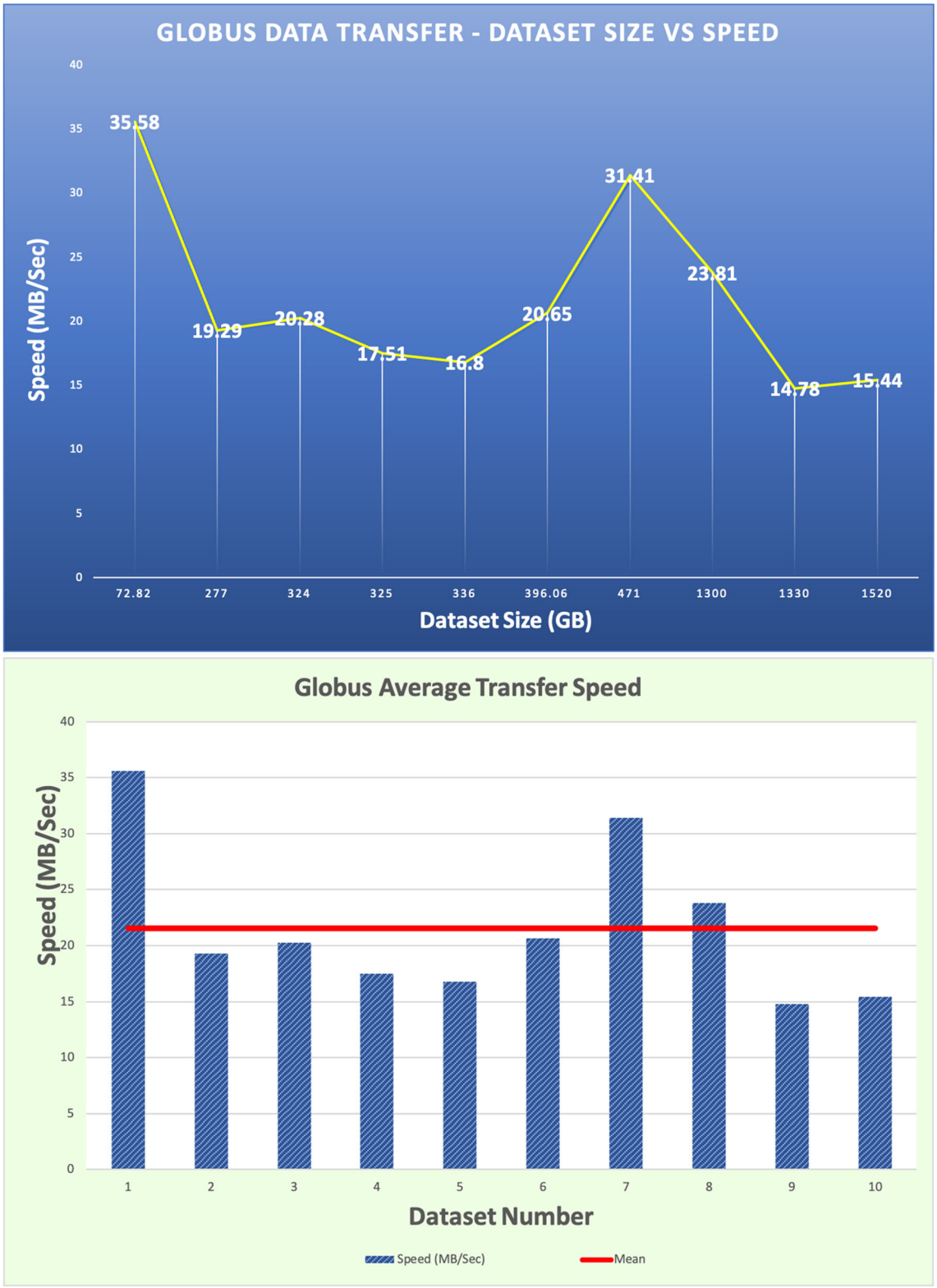

Fig. 4 Evaluation of Globus Transfers (a) Globus Data Transfer - Dataset Size Vs Speed (b) Globus Average Transfer Speed

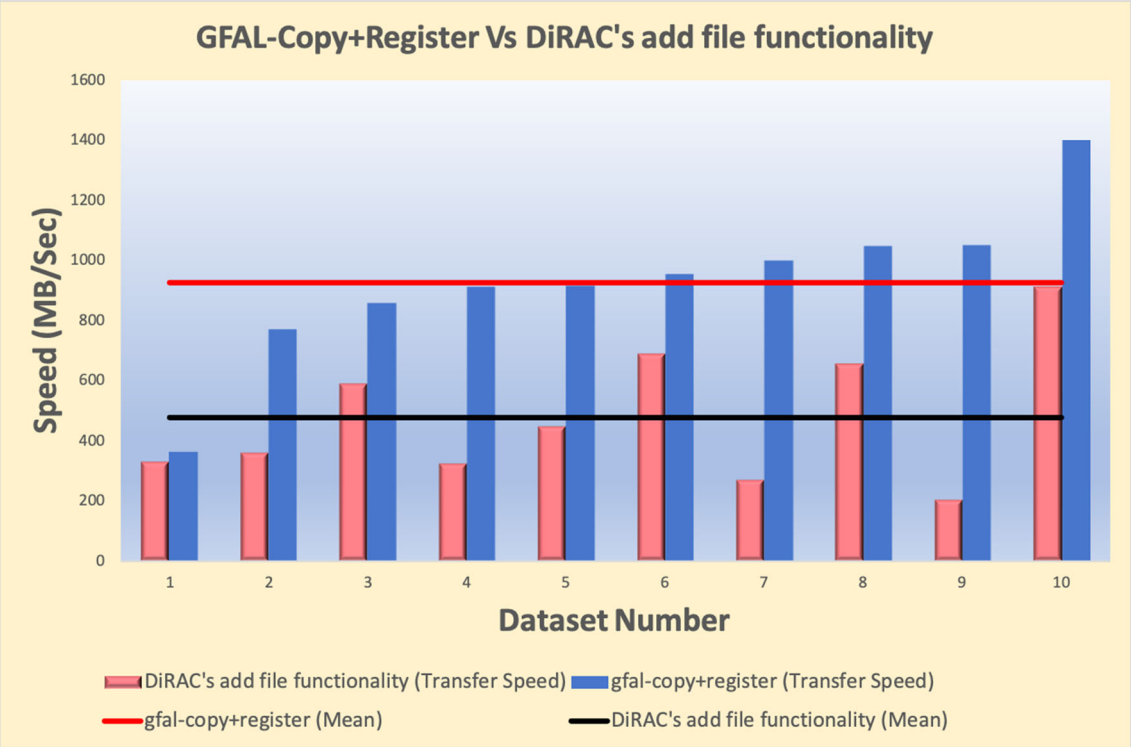

Fig. 5 Comparison of GFAL-Copy+Register Vs DiRAC's add file functionality

are represented in the Table 1. GFAL-Copy's data transfer rate is slightly higher for smaller files and much higher for larger files compared to the DiRAC's Add file functionality.

(c) Efficiency of RapidXfer

Let's say the transfer speed of Globus, GFAL-Copy+register, and RapidXfer as $Speed_{\text{Globus}}$, $Speed_{\text{GFAL-Copy+Register}}$ and $Speed_{\text{RapidXfer}}$ respectively. Further, let's say the size of the dataset be $\text{Data}_{\text{Size}}$ and the total time taken for transferring the $\text{Data}_{\text{Size}}$ for Globus and GFAL-Copy+Register be $\text{Time}_{\text{Globus}}$ and $\text{Time}_{\text{GFAL-Copy+Register}}$ respectively. The Speed of RapidXfer i.e. $\text{Speed}_{\text{RapidXfer}}$ for each dataset is calculated by the following formula (1).

$$Speed_{\text{RapidXfer}} = (Data_{\text{Size}})/(Time_{\text{Globus}} + Time_{\text{GFAL-Copy+Register}}) \quad (1)$$

Table 1 Comparison between GFAL-Copy+Register and DiRAC Add File

Comparison Metric	GFAL-Copy+Register	DiRAC-Add File
Highest transfer rate	1398.52 MB/Sec	909.29 MB/sec
Lowest transfer rate	363.76 MB/Sec	203.89 MB/Sec
Average transfer rate	925.38 MB/Sec	477.57 MB/Sec

Figure 6a shows the transfer time for datasets ranging from 72.82 GB to 1.52 TB arranged from smallest to largest data size. RapidXfer took 35.16 minutes to transfer the smallest dataset 72.82 GB took whereas it took 27.87 hours to transfer the largest dataset 1.52 TB. The transfer rate for RapidXfer is shown in Fig. 6b for datasets ranging from 72.82 GB to 1.52 TB arranged from smallest to largest data size. The Lowest transfer speed 14.57 MB/Sec is recorded for the dataset of size 1.33 TB and the highest transfer speed 34.52 MB/Sec is recorded for the dataset of size 72.82 GB.

Data transfer rates of the 10 testbed datasets fall in the 5 speed ranges i.e. 10.1 - 15 MB/sec, 15.1 - 20 MB/sec, 20.1 - 25 MB/sec, 25.1 - 30 MB/Sec, 30.1 - 35 MB/Sec and they are represented in the Fig. 6c. From this, we can interpret that 50% of the data transfers falls in the speed range between 15.1 MB/sec and 20 MB/sec. Out of the 10 datasets of sizes ranging from 72.82 GB to 1.52 TB taken for evaluation, 70% of the data transfers fall in the speed range between 15.1 MB/Sec and 25 MB/Sec. The average transfer speed of RapidXfer is 20.06 MB/Sec which is shown in Fig. 6d.

6 Conclusion and and future work

In this paper, we have seen the data management challenges and open issues expected for the world's largest telescope Square Kilometre Array (SKA). Due to the unprecedented

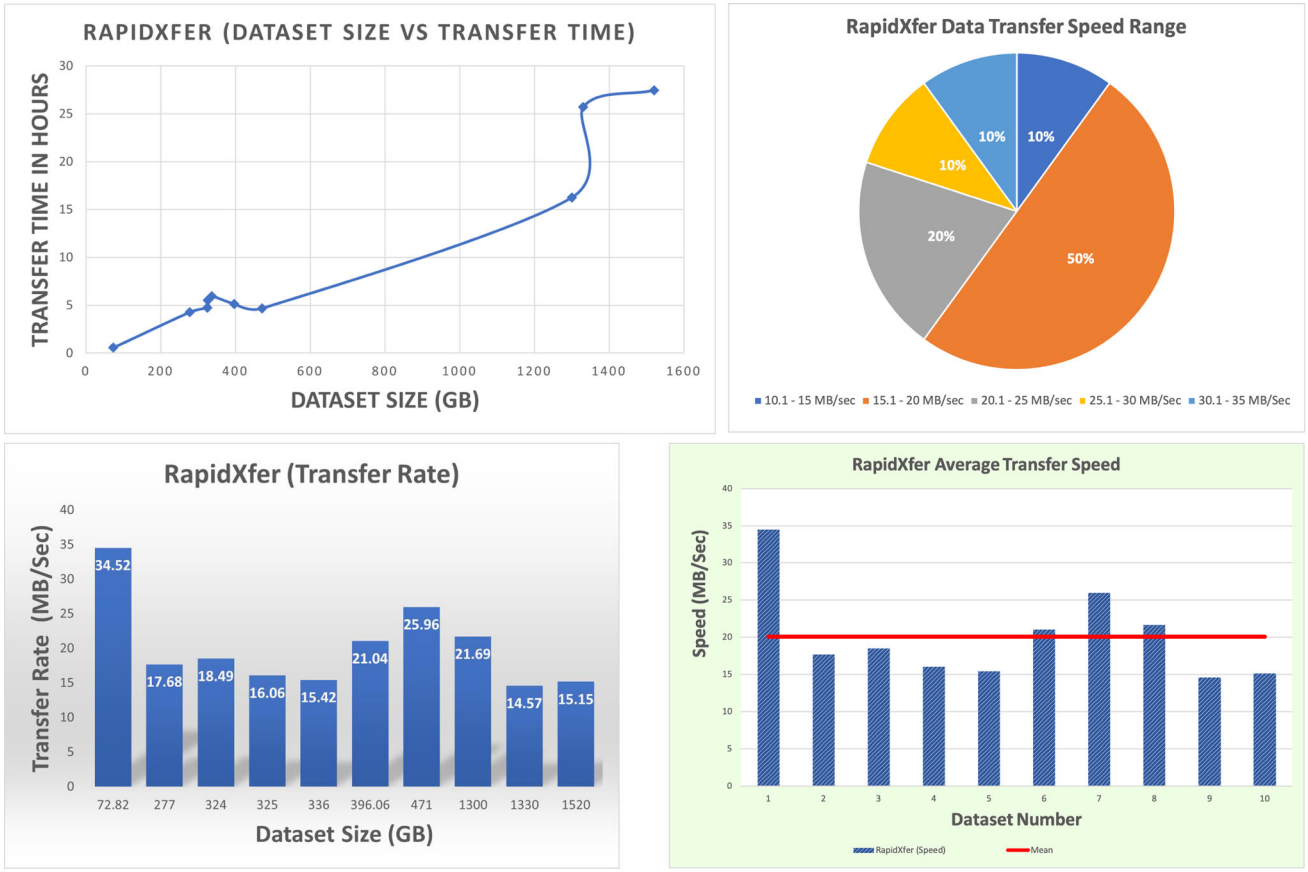

Fig. 6 Evaluation of RapidXfer **a** RapidXfer (Dataset size Vs Transfer Time) **b** RapidXfer (Transfer Speed) **c** RapidXfer Data Transfer Speed Range **d** RapidXfer Average Transfer Speed

amounts of data coming from the telescope, the traditional data transfer methods cannot be used and it is essential to build new data transfer method with higher data transfer speed. This paper focuses on only the MeerKAT telescope data. MeerKAT is one of the precursor telescopes of SKA and the data transfer method discussed in the paper will only be applicable for transfer between South Africa's IDIA site to UK's IRIS grid storage. The proposed data transfer framework 'RapidXfer' takes advantage of Globus and GFAL. It is because of the rigid data management service of IRIS grid storage, the two-step mechanism is needed for the end-to-end data transfer. The implementation of RapidXfer is discussed in the paper. The efficiency of different data transfer tools including `Globus` and `GFAL-Copy+Register` used in `RapidXfer` is evaluated on varied datasets. Further, `GFAL-Copy+Register` is compared with `DiRAC's Add File` functionality to understand the pros and cons of the 2 methods. Current usage of ' GFAL-Copy+Register' is performed without checksum calculation, it is planned to test the efficiency of GFAL-Copy with checksum functionality. It is also worth mentioning that the `RapidXfer` is an interim solution that is currently being used for transferring MeerKAT data to IRIS grid storage. In the future, it is planned to test the performance of `RapidXfer` with even bigger datasets in sync with the Globus updates.

Acknowledgements I would like to thank staff members at SKA and the University of Manchester for providing access to the resources for research.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

Aikema D, Frank B, Simmonds R, Grange Y, Sánchez-Expósito S, Gaudet S, Goliath S (2020) Data Delivery Architecture for MeerKAT and the SKA. In: Ballester P, Ibsen J, Solar M, Shortridge K (eds) Astronomical data analysis software and systems XXVII, astronomical society of the pacific conference series, vol 522. Astronomical Society of the Pacific Conference Series, p 331

Allen B, Bresnahan J, Childers L, Foster I, Kandaswamy G, Kettimuthu R, Kordas J, Link M, Martin S, Pickett K, Tuecke S (2012) Software as a service for data scientists. Commun ACM 55(2):81–88. https://doi.org/10.1145/2076450.2076468

Array SK (2019) Ska regional centre steering committee (srcsc). https://indico.skatelescope.org/event/559/contributions/6225/attachments/5293/7343/SRCSC_ToR_FINAL.pdf

Ayllon AA, Salichos M, Simon MK, Keeble O (2014) FTS3: New Data Movement Service For WLCG. J Phys Conf Ser 513(3):032081. https://doi.org/10.1088/1742-6596/513/3/032081

Barisits M, Beermann T, Berghaus F, Bockelman B, Bogado J, Cameron D, Christidis D, Ciangottini D, Dimitrov G, Elsing M, Garonne V, Girolamo AD, Goossens L, Guan W, Guenther J, Javurek T, Kuhn D, Lassnig M, Lopez F, Magini N, Molfetas A, Nairz A, Ould-Saada F, Prenner S, Serfon C, Stewart G, Vaandering E, Vasileva P, Vigne R, Wegner T (2019) Rucio - scientific data management. CoRR 1902.09857

Bauer D, Colling D, Currie R, Fayer S, Huffman A, Martyniak J, Rand D, Richards A (2015) The GridPP DIRAC project - DIRAC for non-LHC communities. J Phys Conf Ser 664(6):062036. https://doi.org/10.1088/1742-6596/664/6/062036

Britton D, Cass A, Clarke P, Coles J, Colling D, Doyle A, Geddes N, Gordon J, Jones R, Kelsey D, Lloyd S, Middleton R, Patrick G, Sansum R, Pearce S (2009) GridPP: The UK grid for particle physics. Phil Trans Ser A Math Phys Eng Sci 367:2447. https://doi.org/10.1098/rsta.2009.0036

Chard K (2014) Efficient and secure transfer synchronization and sharing of big data. https://www.globus.org/sites/default/files/Efficient_and_Secure_Transfer_Synchroniz.pdf

Chard K, Foster I, Tuecke S (2017) Globus: Research data management as service and platform. 1–5. https://doi.org/10.1145/3093338.3093367

Collaboration CMS, Chatrchyan S, Hmayakyan G, Khachatryan V, Sirunyan AM, Adam W, Bauer T et al (2008) The CMS experiment at the CERN LHC. J Instrum 3(08):S08004. https://doi.org/10.1088/1748-0221/3/08/s08004

Dewdney PE, Hall PJ, Schilizzi RT, Lazio TJLW (2009) The SKA Project. Proc IEEE 97(8):1482

Globus (2021) Data transfer with globus. https://www.globus.org/data-transfer

globus.org (2012) globusa uchicago non-profit service. https://www.globus.org

globus.org (2021) globus docs - faqs: Transfer and sharing. https://docs.globus.org/faq/transfer-sharing/#what_is_the_effective_transfer_rate_reported_by_globus

Haen C, Charpentier P, Frank M, Tsaregorodtsev A (2015) The DIRAC Data Management System and the Gaudi dataset federation. J Phys Conf Ser 664:042025. https://doi.org/10.1088/1742-6596/664/4/042025

Inter-university institute for data intensive astronomy. https://www.idia.ac.za/about-idia/ (2015)

JBC for Astrophysics (2009) The square kilometre array (ska). http://www.jodrellbank.manchester.ac.uk/research/research-centres/ska-project/

Kiryanov A, Ayllon AA, Keeble O (2015) FTS3 / WebFTS? Äì a powerful file transfer service for scientific communities. Procedia Comput Sci 66:670. https://doi.org/10.1016/j.procs.2015.11.076. http://www.sciencedirect.com/science/article/pii/S1877050915034250. 4th International Young Scientist Conference on Computational Science

Liu W, Tieman B, Kettimuthu R, Foster I (2010) A data transfer framework for large-scale science experiments. In: Proceedings of the 19th ACM international symposium on high performance distributed computing, HPDC '10. Association for Computing Machinery, New York, pp 717–724. https://doi.org/10.1145/1851476.1851582

Liu Z, Balaprakash P, Kettimuthu R, Foster I (2017) Explaining wide area data transfer performance. In: Proceedings of the 26th international symposium on high-performance parallel and distributed computing, HPDC '17. Association for Computing Machinery, New York, pp 167–178. https://doi.org/10.1145/3078597.3078605

SARAO (2018) Breakthrough listen to incorporate the meerkat array in its existing search for extraterrestrial signals and technosignatures https://www.sarao.ac.za/media-releases/breakthrough-listen-to--incorporate-the-meerkat-array-in-its-existing-search-for-extraterrestrial-signals-and-technosignatures/

Science TF (2019) Council. Square kilometre array (ska). https://stfc.ukri.org/research/astronomy-and-space-science/astronomy-space-science-programme/ska/

STFC-IRIS (2021) Cutting edge science needs cutting edge digital infrastructure. https://www.iris.ac.uk/

STFC-IRIS (2019) Who are the iris partners? https://www.iris.ac.uk/about-iris/partners/

T.G.C. 2020 (2020) Gridpp: Distributed computing for data-intensive research. https://www.gridpp.ac.uk/

Tsaregorodtsev A, Garonne V, Stokes-Rees I (2004) The GridPP DIRAC project - DIRAC for non-LHC communities. In: Fifth IEEE/ACM international workshop on grid computing, pp 19–25

U. SQUARE KILOMETRE ARRAY (2019) Ska location. https://unitedkingdom.skatelescope.org/ska-location/

W.L.C. Grid (2006) Worldwide lhc computing grid. https://wlcg-public.web.cern.ch/