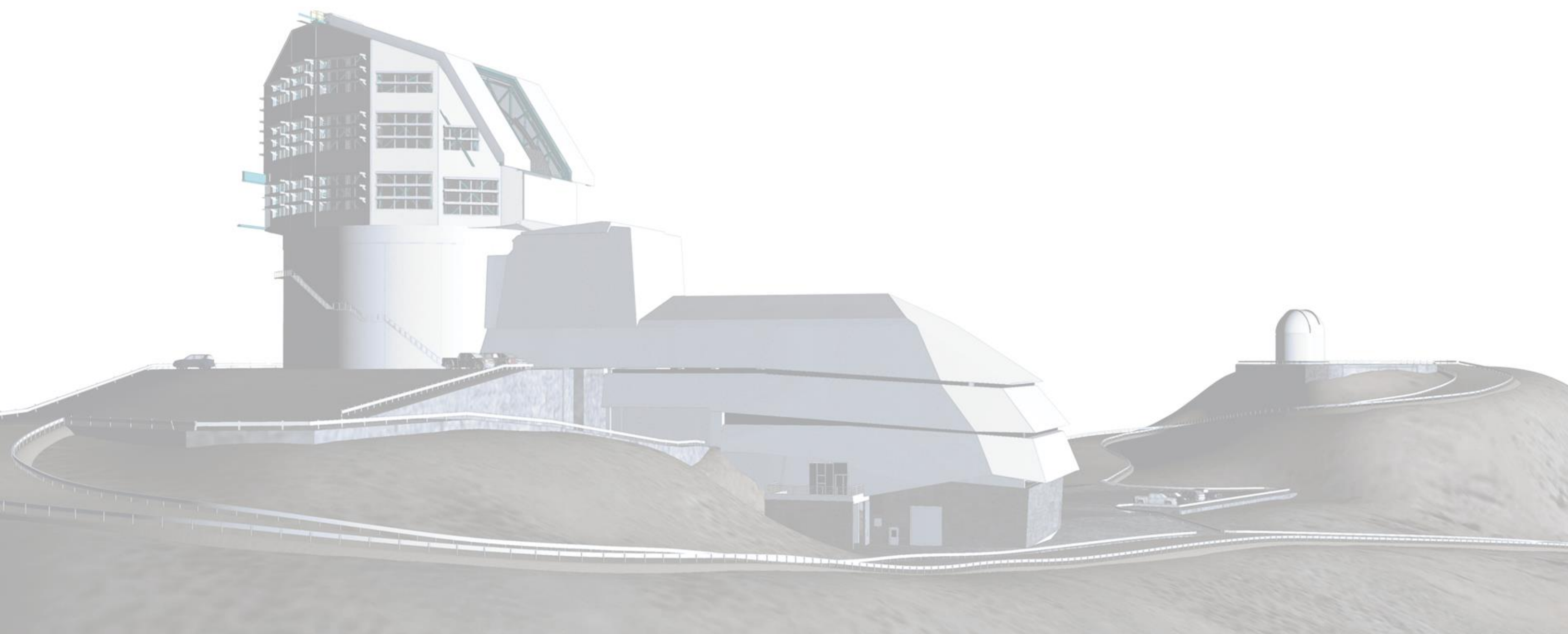
Observatory Network Design

Ron Lambert, Jeff Kantor, Mike Huffer, Chip Cox, Paul Wefel, Matt Kollross, Albert Astudillo, Sandra Jaque, Mark Foster, Richard Hughes-Jones, Stuart McLellan

LSE-78 (rel6.0)

Latest Revision Date: September 23, 2021



Change Record

Release	Date	Description	Owner Name
1	3/18/2009	Initial version as Document-7354	Ron Lambert and Jeff Kantor
1.2	9/10/2010	Added US national network information	Jeff Kantor
1.3	9/22/2010	Minor Updates	German Schumacher
1.4	6/3/2011	Summit network updates	German Schumacher
2	6/16/2011	Elevated to LSE handle and updated with latest information from MREFC update and 10-40-80 plan	Jeff Kantor
2.1	7/8/2011	Update for Directors Review and PDR; added AmLight figure	Jeff Kantor
2.2	8/5/2011	Updated Summit diagram	Jeff Kantor
2.3	2/25/2013	Adjusted for guaranteed 10 Gbps (vs 10/40/80), updates to MIA - NCSA link, and AP move to NCSA	Ron Lambert, Jeff Kantor
2.4	6/7/13	Updated international text and diagrams. Reorganized text into sections for clarity. Numerous edits to reflect latest network	Jeff Kantor, Chip Cox

		baseline.	
2.5	9/6/13	Updated for increase Mountain – Base network bandwidth to 200 Gbps	Jeff Kantor
	10/4/2013	Changes approved 10/3/2013 – LCR-139	R. McKercher
2.5.1	11/1/2013	Corrected Figure 2 to 10M alerts/night	G. Dubois-Felsmann
3.0	3/9/2015	Mtn - Base and La Serena - Santiago	Ron Lambert
4.0	8/25/2015	Implementation of LCR-375; no content change since Post-INAR iteration (version-30904)	R. McKercher
4.1	12/15/2015	Mtn-Base diagrams and fiber updates.	R. Lambert
4.1.1	12/17/2015	Post-CCB-meeting edits, Figure 5 replaced.	Ron Lambert
4.1.2	12/18/2015	"architecture… may change" sentence removed; "to be updated" markings for Fig. 5 removed	Gregory Dubois-Felsmann
4.2	1/14/2016	New Figure 5 received from Ron Lambert	Ron Lambert / gpdf
4.2.1	1/25/2016	Figure 11 rescaled to fit page, no content change.	Gregory Dubois-Felsmann
4.2.1	2/1/2016	Implementation of changes approved via LCR-385	R. McKercher

4.3	1/19/2017	Updated to reflect developments in all networks, especially Summit Network	J. Kantor
	4/27/2017	Implementation of LCR-849 and release 4.3	R. McKercher
5.0	5/5/20	As-built update to diagrams, text Change LSST to Rubin Observatory Reference to Rubin Observatory Networks Technical Document Additional information on NCSA to CC-IN2P3 link Remove references to science image transfer Misc. editorial updates (e.g. use of tense)	J. Kantor R. Lambert A. Astudillo J. Bezerra M. Kollross
5.1	6/26/20	Incorporated comments from Networks Pre-Verification Review: Eliminate reference to NLR Eliminate reference to dynamic l2 at DAC	J. Kantor J. Bezerra
6.0	2021-09-23	Updated with USDF at SLAC and SLAC – IN2P3 Connections. Updated ESnet sections. Implemented via LCR-2798	J. Kantor, M. Foster, R Hughes-Jones, S. McLellan, P. Wefel

Table of Contents

Observatory Network Design

1 Introduction

This document provides the top-level design for the Rubin Observatory long-haul networks connecting the various operational sites in Chile, the United States, and Europe. The detailed design for these networks is captured in the Rubin Observatory Networks Technical Document.

The Rubin Observatory is distributed over seven sites: the Summit Site, the Base Site, the Archive Site, the Headquarters Site, the Camera Site, the French Site, and the United Kingdom Site. Figure 1 shows the location, features, and function of each site. Each site hosts one or more Rubin Observatory Centers and Facilities, which are purpose-specific, staffed, operating entities within the Rubin Observatory. Each Center/Facility has personnel and equipment housed there in order to perform its assigned functions.

While the sites are geographically distributed, they are all functionally integrated. Dedicated, AURA-owned fiber optic lines connect the Summit and Base Sites, while the others are connected through fibers on carrier-owned networks. Control functions are distributed for operational efficiency and to provide robust, reliable, safe operation.

During operations, the Rubin Observatory Project will broadcast near real-time transient alerts within one minute of image collection. Digitally processed images and feature catalogs will be available to the U.S. and Chilean communities via public networks within 24 hours. Full data product releases are scheduled every year, with two planned releases with data from the first year of operations.

The Mountain Summit Site is located on Cerro Pachón, Chile, at the telescope site and performs readout of the camera. The Base Site is located at the AURA compound in La Serena, Chile, and provides a Data Access Center (DAC) for the Chilean astronomical community. The Archive Site is located at the SLAC National Accelerator Facility (SLAC) in Menlo Park, California. It provides transient alert processing, calibration data products processing, and data release processing, as well as complete data product archiving. The Archive Site houses the Data Facility and a DAC for the U.S. community. The French Site is located at the CC-IN2P3 computing center in Lyon France, and provides one half of the processing capacity for the annual Data Releases, via an approved Memorandum of Agreement. The UK site is located at the Royal Observatory Edinburgh and provides a quarter of the processing capacity. The Headquarters Site is currently in Tucson, Arizona and provides overall observatory management, science operations, and EPO functions.

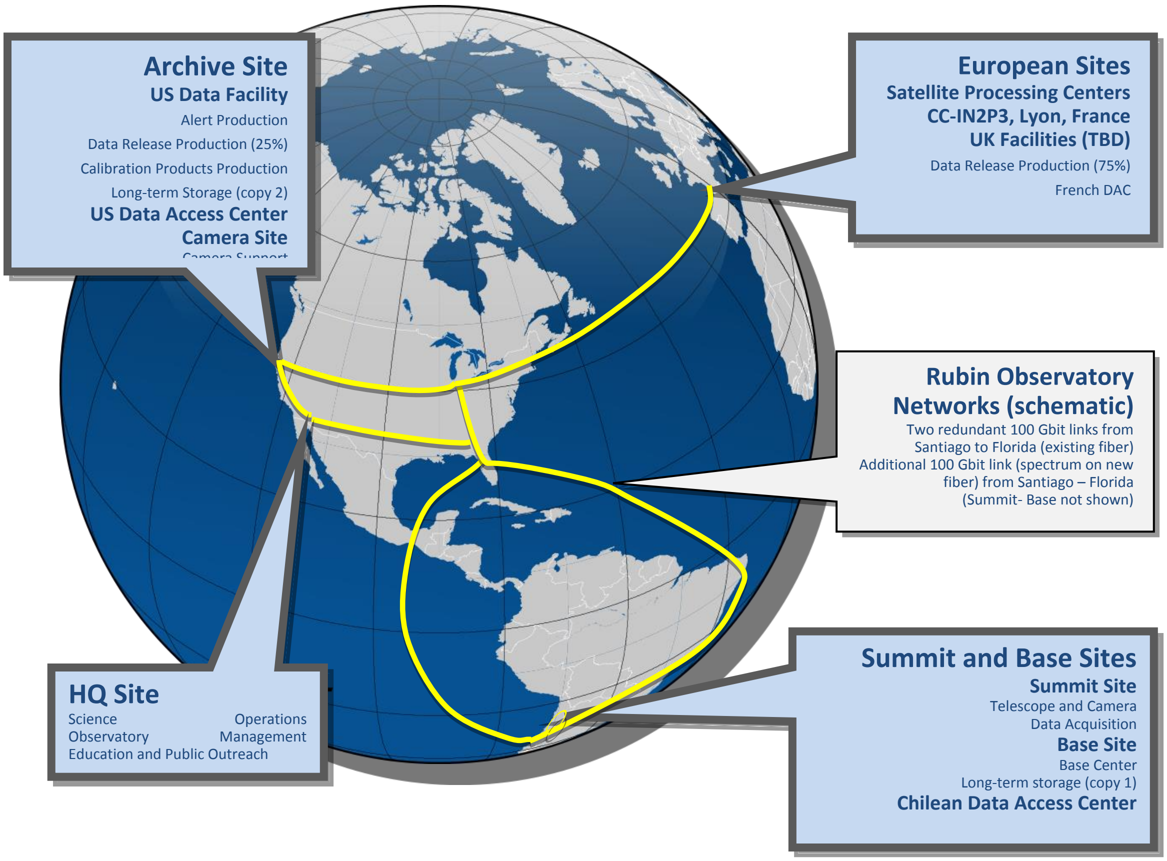


Figure 1 Vera C. Rubin Observatory Sites and Networks

2 Vera C. Rubin Observatory Network Requirements and Data Flows

2.1 Network Requirements

The requirements most directly driving Rubin Observatory Network design specify the capabilities to transfer data and control information within and between the subsystems of Rubin Observatory (Telescope and Site, Camera, Data Management, Education and Public Outreach), as well as providing access to the data to scientists and non-scientist users.

Rubin Observatory has a wide range of time scales in these requirements, including

- Order of milliseconds in the Observatory Control System (OCS) and control systems of each Rubin Observatory subsystem (TCS, CCS, DMCS),
- Order of seconds in acquiring and processing the raw image stream from the Camera through Data Management in order to create transient alerts,
- Order of days in archiving the raw image stream and other metadata for subsequent use in the production of astronomical catalogs and other data release products,
- Order of weeks to months in producing and deploying data necessary for calibration of the Telescope and Camera, and
- Order of years in producing and deploying Data Release of Rubin Observatory Legacy Survey of Space and Time (LSST) processed images and catalogs.

In addition, there are large geographical distances between the various Rubin Observatory sites, which imply non-trivial latency in certain data transfers.

The above requirements are documented in Rubin Observatory System Specifications (OSS, LSE-30), and they are flowed down to Rubin Observatory subsystem requirements documents, e.g. Data Management System Requirements (LSE-61).

There are also data transfer requirements associated with supporting user access in DACs in Chile and the United States, and in the Education and Public Outreach Center (EPOC) in the United States. In the case of Chile, these requirements are derived from Memoranda of Agreement (MOA) between Rubin Observatory and AURA [AURA Rubin Observatory MOA], and between AURA and Chile [AURA Chile MOA] that require providing network infrastructure and a DAC in Chile in return for authorization to site Rubin Observatory in that country.

All of these requirements and MOA drive the need to provide high-speed networking between the various Rubin Observatory sites and centers, and in turn have driven the network design.

Early in the design process, we evaluated the alternative of shipping media for the less time-constrained requirements, i.e. where the full preparation, packing, shipping, unpacking, and loading time, estimated at 3 - 5 days, was shorter than the required time frame. We realized that we could effectively use shipped media for the annual deployment of Data Releases, since we have annual storage hardware refreshes being shipped in those time frames. In the event that the anticipated bandwidth between the Archive and Base Sites is not realized, or has an extremely long extended outage, this remains an option for transferring Data Releases to the Base Site. However, given the need to provide high-speed networks for the more stringent time-constrained requirements, the data release and calibration data transfer will also be accomplished on these networks.

Rubin Observatory has invested in network infrastructure (fibers, equipment) that will guarantee adequate network links from Cerro Pachon to La Serena to North America during Commissioning and Operations. In addition, the French IN2P3 institute is investing in connectivity between the Archive Site and the Satellite Processing Site in Lyon, France. The costs, bandwidths, and utilization of the network links by Fiscal Year are specified in LDM-142 Network Sizing Model.

The La Serena to SLAC network segments are specified in every major segment as at least 2 diverse paths. From Santiago to Florida one segment transits the Pacific coast of South America and two others via the Atlantic coast, including new submarine and terrestrial cables on which Rubin Observatory will have spectrum. Each path independently has sufficient bandwidth to support all science requirements including transient alert production.

2.2 Data Flows

2.2.1 Normal Operations – Alert Processing

LSST will discover large numbers of astrophysical objects that move across the sky, change in brightness, or appear as transients. LSST's real-time Prompt Processing pipelines will identify such detections using image differencing and report them using world-public *alerts* issued within 60 seconds of the shutter closure after each visit. LSST expects to produce up to about ten million of these alerts nightly. The resulting *alert stream* will be large—more than 1 TB nightly.

To analyze these alerts, astronomers will rely on third party *community brokers*, software systems that receive the full LSST alert stream and provide additional information to refine the selection of events of interest. LDM-612 ("Plans and Policies for LSST Alert Distribution") outlines the overall alert distribution system and describes the process by which LSST will choose community brokers.

LSST is required to be able to support a full alert stream to at least five community alert brokers (DMS-REQ-0391). To transmit alerts from the Data Facility to community brokers within the required latency, sizing calculations suggest that an instantaneous outbound bandwidth of up to 5.4 Gbps is required. DMTN-102 details these calculations.

Terabyte-sized volumes of Rubin Observatory data and data products must be moved each night and day from the Summit to the Base Center in Chile, from the Base Center to the Data Facility and Data Access Centers, and finally from those centers to the users when necessary, as depicted in Figure 2.

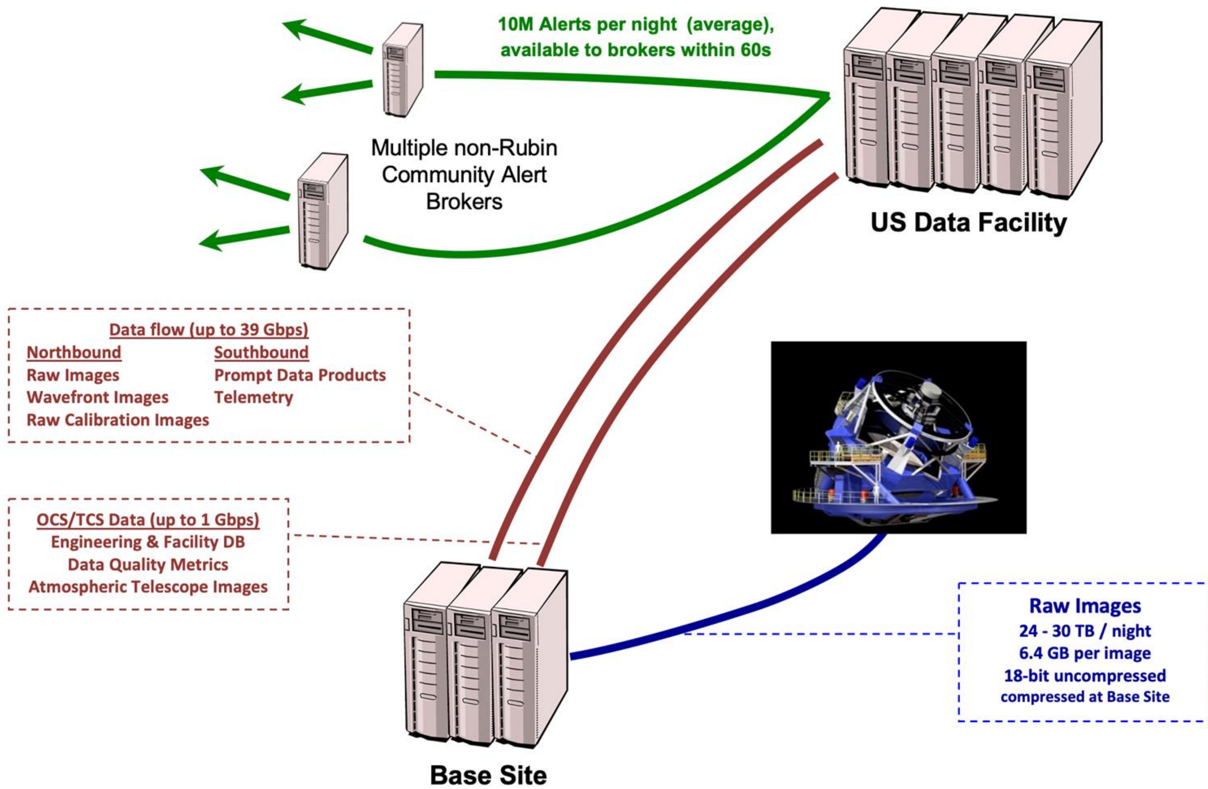


Figure 2 Nightly International Data Flows

2.2.2 Normal Operations –Data Release Processing

Raw image data is transferred from the US to the UK and French Data Facilities on a routine basis as the Survey proceeds. It remains to be determined if all the images will be sent to France from the US Data Facility and, from there, 25% of the images to the UK, or if the US would send the 25% directly to the UK. After producing their shares of the DRP, each Data Facility will update the others, with the US and France facilities having complete copies. After single image processing is complete, the CoAdd step is performed and resulting data are distributed. Object catalogs are also distributed at this time.

The Data Release is transferred to the US DAC via networks internal to the Data Facility, and subject to capacity in La Serena (see LDM-572 Chilean Data Access Center), to the Chilean DAC via the international network. Media transfer with an annual hardware upgrade is a secondary option. There is also opportunistic transfer via the network of sky templates, co-added images, calibration data, and catalog updates from the Data Facility to the Base Center at more frequent intervals.

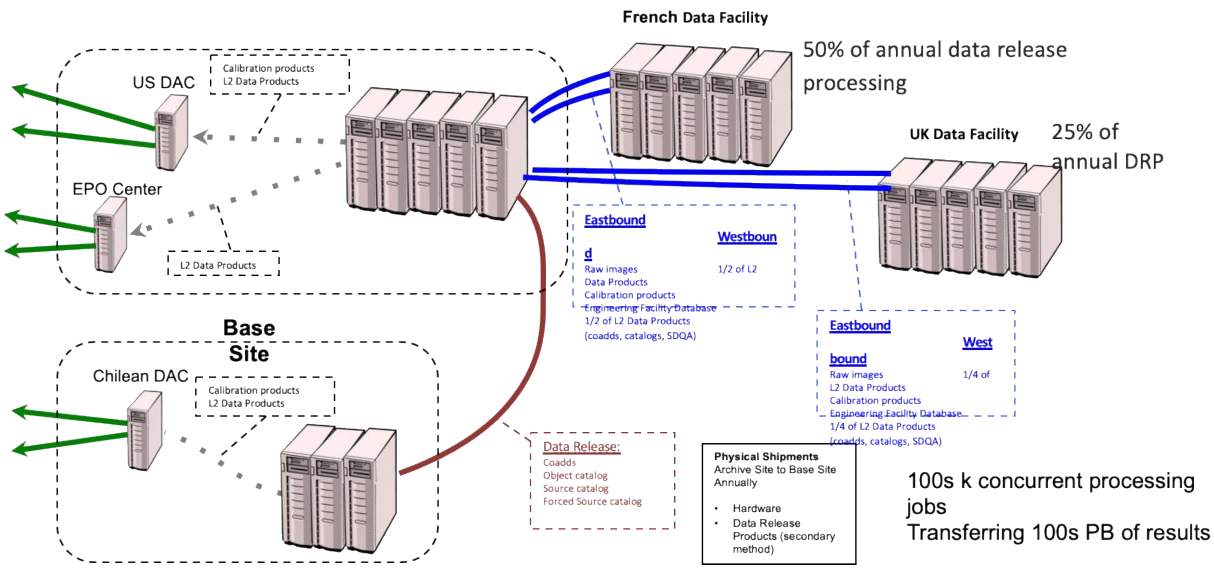


Figure 3 Rubin Observatory Data Release Processing Data Flow Diagram

Figure 17 also shows the primary and secondary access connections between GÉANT and Janet and GÉANT and RENATER which are sited at different physical locations in each country and currently (May 2021) 100 Gigabit. These divers connections help ensure reliable access to the ESnet Trans-Atlantic links and hence provide the required performance and backup paths for the end-to-end connectivity between SLAC and the European data processing centres. The core or backbone network infrastructure in each NREN use multiple 100 Gigabit links. Data exchanges between the European data processing centres and DACs will be carried over and Janet and GÉANT and RENATER

2.2.3 Sizing and Architecture

2.2.3.1 Summit to Base Link

The key driving requirements for the Summit to Base communications are the bandwidth and reliability required to transfer the image data and associated EFD meta-data for alert processing, (including transferring the data to the Base to Archive network), and to handle OCS command and control traffic.

Rubin Observatory installed a fiber pair, and utilizing DWDM technology, provisioned three high bandwidth Lambdas with end nodes that are owned by Rubin Observatory. This capacity is nominally divided into flows for image data transfer and a separate flow for OCS command and control data, but these allocations may change based on observatory needs. Because this link closely follows the path of existing CTIO networks, REUNA will operate, and Rubin Observatory will manage and maintain this mission-critical link, providing the reliability and availability necessary to run the Summit to Base infrastructure for the Rubin Observatory.

2.2.3.2 Base to Archive Link

End-to-end analysis has shown that in order to keep up with the observing cadence, a minimum of 40 Gbps is required and this is the design baseline. This minimum required peak bandwidth from the Base Center to the Data Facility is established by the need to transfer the raw science images for transient alert processing. The peak requirement for annual data release transfer from the Data Facility to the Base Center is less, as this can occur over the period of a month or more.

In order for the Data Facility to process transient alerts within 60 seconds of science image readout from the camera electronics, we assumed 7s of this time budget is allocated to data transfer, including compression/decompression. We assumed 2s of that budget is in the Summit to Base transfer, and the other 5s is in the Base to Archive transfer and compression/ decompression. In the event encryption/decryption is required on the transfer, the time budget for those functions will come from the overall 60s prompt processing budget, while the transfer budget will remain as is.

Analysis and testing show that the transfer time from the Base to the Archive can be accomplished within the required time frame on tcp-based 4 x 10 and 100 Gbps networks, even when considering packet loss and latency in the network at much higher levels than are currently being experienced. This is feasible with Virtual Local Area Network (VLAN) protocols to ensure dedicated bandwidth, use of jumbo frames, and configuration of TCP congestion control to optimize utilization in Rubin Observatory traffic patterns. Additional analysis and testing verified that the effect of protocol translation overhead is minimal and does not affect the design. [Freemon2013].

Rubin Observatory has partnered with NSF's Optical-Infrared Astronomy Research Laboratory (NOIRLab) and coordinated with other South American observatories and RENs via the South American Astronomy Coordination Committee (SAACC) to leverage and expand existing NSF-funded and South American national network initiatives that have allowed increasing bandwidth capacity between Chile and the United States at the same funding level.

This network infrastructure is path diverse from Santiago flowing either side of South America (2 Atlantic, 1 Pacific). Rubin Observatory-available bandwidth on each coast of South America will be more than adequate for Rubin Observatory requirements. Provisions for "bursting" will be available utilizing the two diverse links from Santiago to the United States. The requirement calls for no other traffic to be present on the image data stream link during Observing hours. Moreover, the ultimate goal is to establish the network from La Serena to

Miami, and SLAC over the east coast of South America link, for the sole use of Rubin Observatory traffic. This is currently implemented by AmLight, REUNA, RNP, ESnet and Rubin Observatory personnel and by the end of construction, will exist in all segments except for Boca Raton to Atlanta, which will be implemented early in Operations.

The Rubin Observatory networks presume continued investment and development in these networks by the United States, Chile, and Brazil. Contingency plans exist in case those investments are not sustained and only the minimum bandwidth is available.

2.2.3.3 Archive to DAC/EPOC to User Links

Within the United States and Chile, Rubin Observatory must provide open access to Rubin Observatory data—not only to the scientific community, but also to diverse segments of non-specialist users whose curiosity may be triggered by Rubin Observatory educational and public outreach programs. Rubin Observatory will provide DACs in Chile and the US co-located with the Base Center and Data Facility so data transfers between each site will be via local internal networks with 100 Gbps or higher bandwidth. Rubin Observatory will also provide EPO portals for educational and outreach efforts, hosted in the cloud. User access to the DACs will be via public and Research and Education Network (REN) connections (e.g. The Internet and Internet2, XSEDE, ESnet), and the aggregate bandwidth will be limited only by the connectivity of the hosting and using institutions. In cases of stand-alone DACs or science centers funded outside the project, the entity developing and operating the center will be responsible for providing network connectivity to Rubin Observatory Data Facility to enable data transfer.

2.3 Reliability and Availability

The above paths must also be reliable to avoid data transfer bottlenecks, so path diversity and redundancy are required where economically feasible, and there must be spare capacity on the order of the normal bandwidth to "catch up" in the event that a failure or slow-down occurs.

The above designs are targeted to achieve 98% or higher mean data transfer reliability, which is especially important for transient alert processing and requires higher than 98% network availability. Use of the networks in construction and commissioning, as well as the historical data suggest that this is achievable.

Table 1 Chile and International Link Reliability

Link	**Years in operation (REUNA/AMLIGHT)**	**Cuts/outages**	**Longest Outage**	**Mean Time to Repair**	**Mean Availability**
La Serena – Santiago	10*	multiple	4.5 hrs (27 hrs**)	4.5 hrs	99.5%
Santiago – Sao Paolo	5	multiple	23 hrs	5 hours	99.5%
Sao Paolo – Miami (GX))	8	multiple	34.5 days		99.9%
Sao Paolo – Miami (LAN)	8	multiple	22 hours		99.9%
Santiago – Panama - Miami	8	multiple	22 hours		99.9%

2.3.1 Failure Scenarios on the Summit to Base Link

The only link in the network without path diversity is the Summit to Base link. A tradeoff study was conducted to assess the feasibility of path diversity. The result of this study was that the single path was sufficient to achieve the overall reliability necessary, and that a second path would not be a cost-effective increase in reliability. A second path remains a possible up-scope should funds be available.

In addition, tradeoffs were conducted assessing whether to install the fiber on poles or underground, and various maintenance scenarios were analyzed. The result was to employ a combination of poles and underground for 3 segments in this link. [Rubin Observatory11811]

Finally, the size of the Summit buffer is up to 8 nights of observing (depending on specific observing cadence/programs). In the unlikely event that an outage lasts longer than this, provisions for transferring media down the mountain have been incorporated into the system design.

2.3.2 Failure Scenarios on the Base to Archive Links

This analysis summarizes the tradeoffs associated with operations scenarios that involve failure of one or both of the international links. For this analysis, we use the term Pacific and Atlantic when referring to physical cables, and *Science* and *Secondary* when referring to the data being carried, though various other terms may be used in other sections of this document.

Note that in failure scenarios, Rubin Observatory operations traffic gets priority over non-Rubin Observatory AURA traffic, and both get priority over any other traffic.

2.3.2.1 Atlantic Cable Failure

In normal operations, the science data is routed on the primary Atlantic cable (spectrum). The operational data is routed on the secondary Atlantic cable (managed ring). If either

Atlantic cable fails all flow from that cable is automatically and immediately rerouted to the Pacific cable and supersedes all other network uses. This maintains operations and meets the science requirements. When the Atlantic cable is returned to service it can now be used for non-time critical data. Note that the likelihood of simultaneous failures on both primary and secondary Atlantic cables or simultaneous failures on both Atlantic and Pacific cables is deemed too low to warrant detailed consideration, but in this event, data buffering and subsequent catch-up will occur.

2.3.2.2 Pacific Cable Failure

A Pacific cable failure does not affect critical Rubin Observatory science operations directly but interferes with other non-critical data transfers. This does not have an impact on operational requirements.

2.3.2.3 Cables Repaired

When links are returned to service the data flow is normal on the Atlantic cables, and the Pacific cable may be placed into “catch-up” mode. Images that have not yet arrived at the Data Facility may be transmitted on the Pacific cable, whether or not those images are already in transit to the Data Facility on physical media. The entire network returns to its pre-failure usage as soon as all buffered images have been transmitted to the Data Facility.

With respect to shipping media, a separate analysis has been performed that assess whether the media should be spinning disk, flash, solid-state disk, or tape. Several of these options are feasible, and the current baseline is solid-state disk, but we will continue to monitor technology/cost trends. [Rubin Observatory14313]

3 Mountain Summit Infrastructure

The Summit Computer Room is located at the Summit Site, which is at 2700 meters elevation. There are three main elements at the Summit: the Camera Control System (CCS), the Camera Data System (CDS) and Telescope Control System/Observatory Control System (TCS/OCS) with a minimal presence of the Data Management System (DMS). Each of these systems has been developed independently and relies on the local area network infrastructure for interconnections. The network facilitates data flow between the elements and transfers images and other data to and from the Base Site.

The image data from the Camera system readout destined for the Archive Site are transferred via dedicated DWDM ports and do not traverse the Summit LAN but feed directly into the DWDM client-side interface. Other applications and services may use the Summit LAN to acquire images as well. This is a baseline design to enable the data to be available at the Base Center and the Data Facility in close to real time. The Camera system also outputs a complete image for real time display and other monitoring and diagnostic operations. The Camera will store eight days of raw image data in the event that there is a disruption in the link to the Base or Archive Sites and for queries of recent data.

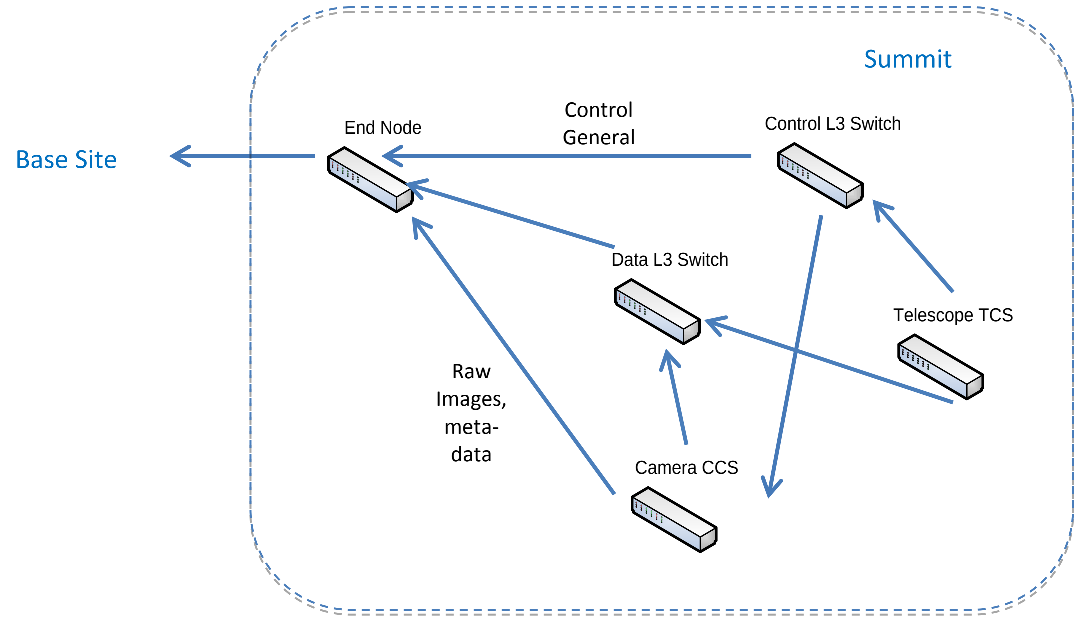


Figure 4 Control and Data Network Diagram

As indicated in Figure 4, lines run over separate networks in order to decrease latency in the system. We maintain separate subnet/VLANs between the different entities in order to impose levels of security with access control lists.

4 Summit to Base Network

The distance from the summit to the base is 55 km point to point or 90 km via the public highway and Observatory access roads. The plan below shows the routing.

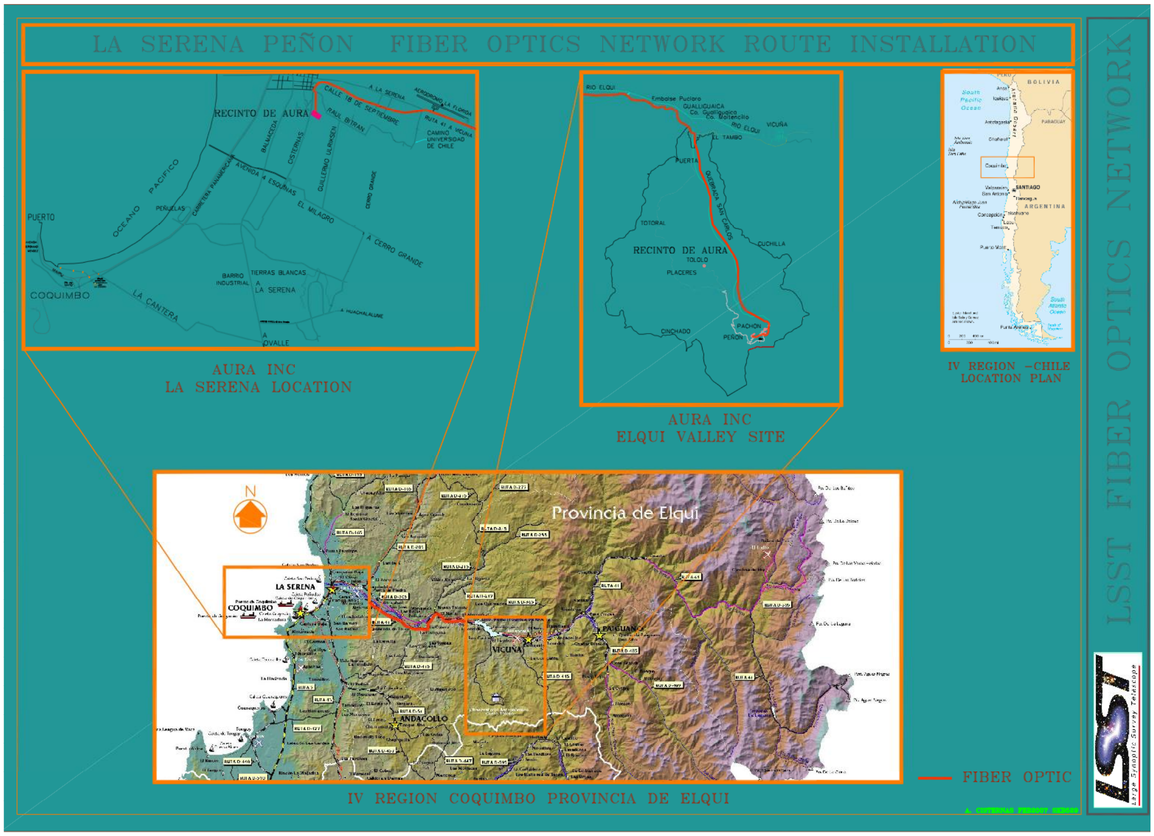


Figure 5 La Serena – Cerro Pachon Fiber Optics Network Route

The geographical diagram above shows the physical routing of the network that has been installed for Rubin Observatory and AURA traffic from summits to base during the first years of Rubin Observatory construction. Note that the route the La Serena fiber takes is not direct from Colina el Pino to the Summit. It goes from Base to OC Telefónica and then to the Summit. That level of detail was omitted in the graphic.

The Summit to Base fiber cable link consists of two distinct portions installed by REUNA/TELEFONICA: Part A La Serena to Gatehouse and Part B Gatehouse to C. Pachón.

Part A La Serena to the AURA Gatehouse in Vicuna

As shown below in Figure 6, there are two pairs of dark fibers provided by Telefonica on the public highway, one pair of which carries only Rubin Observatory traffic and the second pair AURA traffic. These fibers are provided by Telefonica to AURA through an 18-year Irrevocable and Exclusive Use (IEU) and are part of a bundle utilized by the three major telecoms in Chile. Most of the 90Km span is an aerial installation. On the public highway REUNA guarantees 4-6 hour mean time to repair, and on the AURA property 16 hours in daylight.

Figure 6 La Serena to the AURA Gatehouse in Vicuna

Part B Gate to Cerro Pachon

Figure 7 shows the section on AURA property where there is a 12-pair fiber bundle spliced into the Telefonica provided filament fibers in a manner that is consistent with our use and offers spare fibers for redundancy. The cable in fig.8 (blue line) is installed on the electrical posts direct to C. Pachon with a bifurcation at the water pump in the San Carlos valley.

Figure 7 Fiber Path from Gatehouse to Summits

Cable is installed to jumper C. Tololo's traffic over to C Pachon as shown in Figure 8.

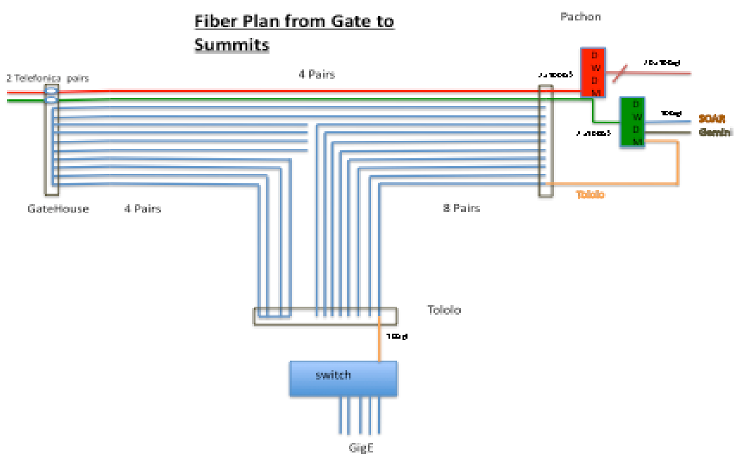


Figure 8 Fiber Plan from Gate to Summits

Equipment for Base to Summit

As shown in Figure 9, the DWDM equipment purchased has the capacity to supply 200G Lambdas. Rubin Observatory provisioned three Lambdas and AURA provisioned one Lambda. Rubin Observatory has the capability via an optical protection circuit to failover from two of the lambdas to the AURA fiber in the event of Rubin Observatory equipment or problems with their fiber pair to La Serena. This is termed a collapsed ring. The redundant link between the AURA and Rubin Observatory DWDM equipment on Cerro Pachón is transported over a 100GigE interface.

Note this is not to indicate that there will be a redundant fiber ring but provides fail-over in the case of an electronic/optical failure or a dirty fiber pair at the local site. It should be noted at this time that redundancy will not be offered for AURA traffic in the Rubin Observatory circuits.

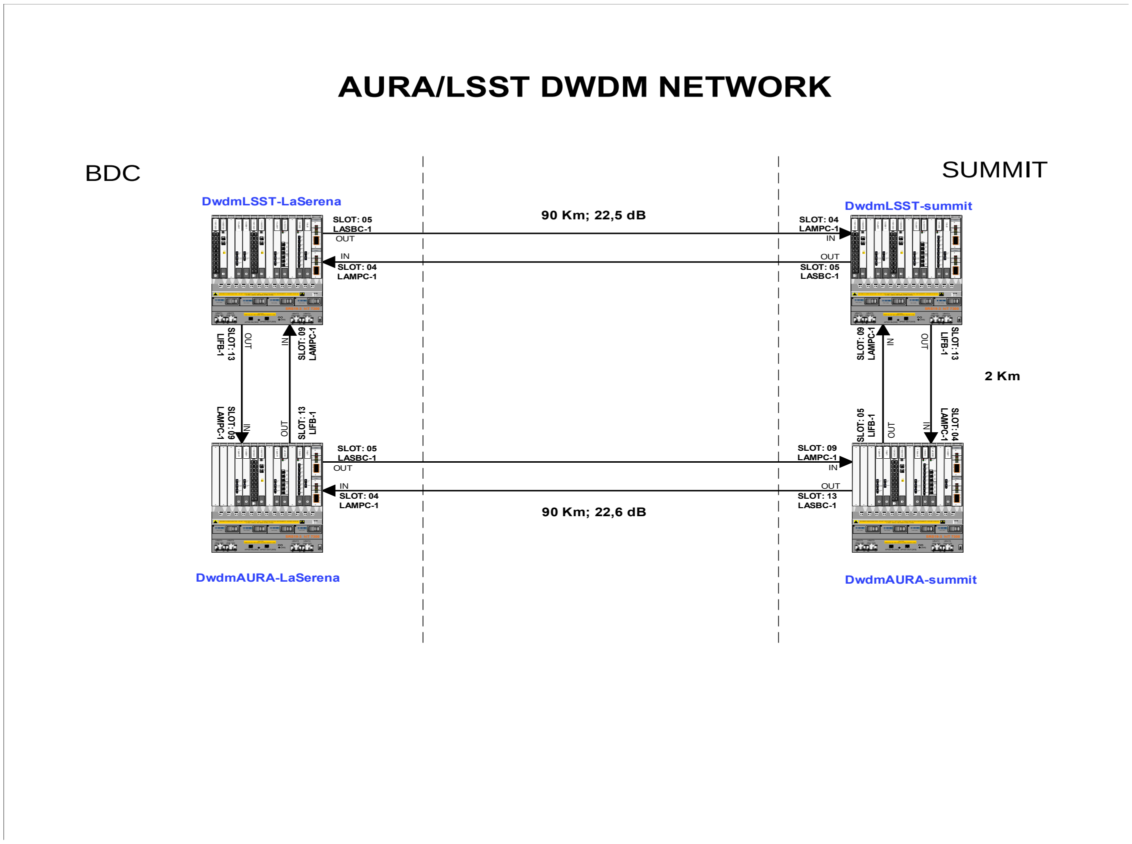


Figure 9 Data flow

As shown in Figure 10, the DWDM chassis houses 40x10GigE modules at layer 2 ethernet on the client-side interfaces for the camera and 2x100GigE interfaces for the Control traffic. On Cerro Pachon the Rubin Observatory chassis physically resides in the Rubin Observatory telescope computer room and the AURA chassis in a shared infrastructure building (caseta). In La Serena the two chassis reside in adjacent racks the Rubin Observatory/AURA data center.

The RFP process took place over a year and the selected DWDM vendor was Coriant (now part of Infinera). A subcontractor Raylex installed the equipment in situ, and provides services and guarantees advertised bandwidths and throughput. REUNA provides operational oversight and support.

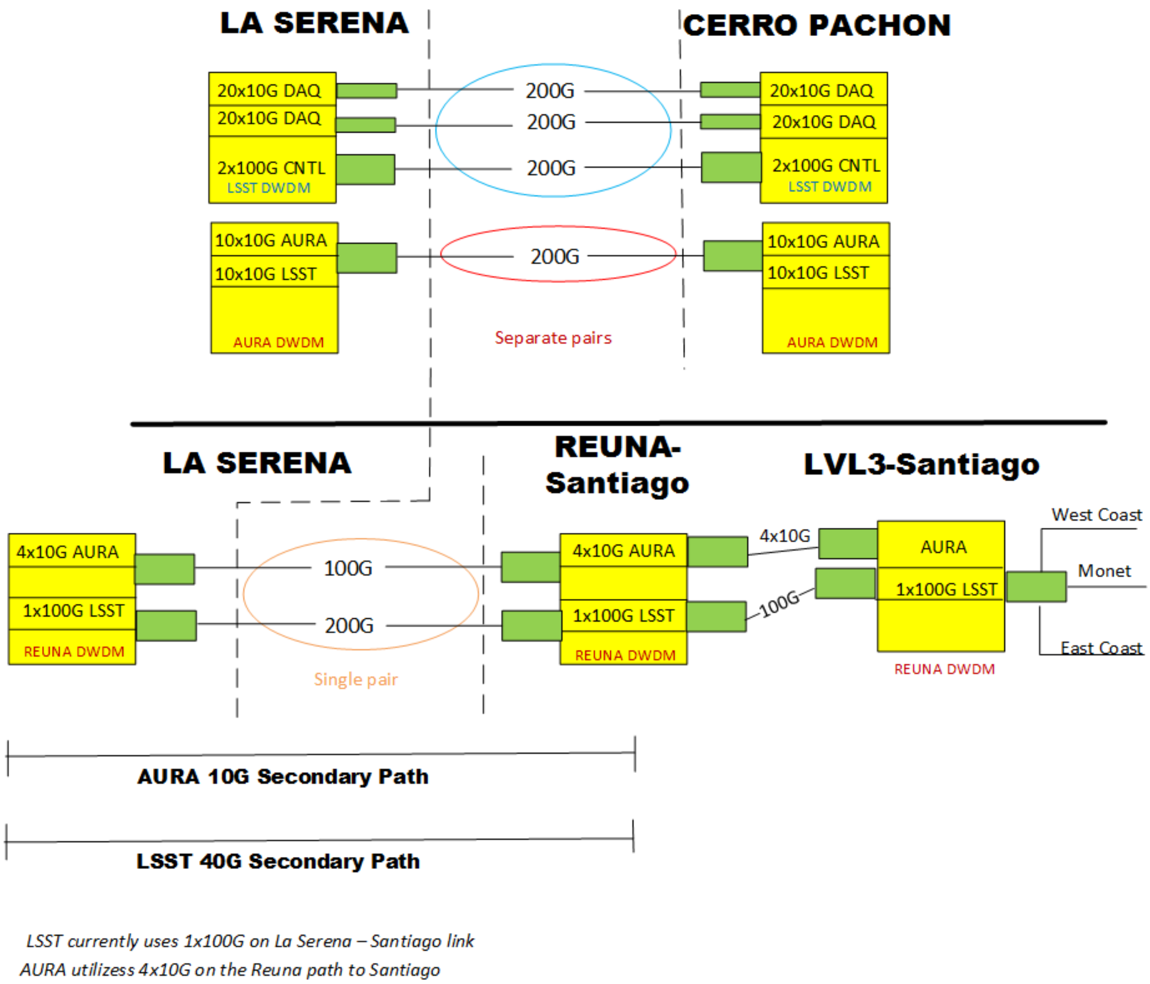


Figure 10 Network End Equipment and Locations

Table 3 shows typical latencies between Cerro Pachón and La Serena across the DWDM and Summit – Base fiber.

Table 2 Typical latency from summit to base

CP DWDM equipment	25uS
Fiber	432uS
LS DWDM equipment (Depending on vendor)	25uS
Total	~480uS

5 Base Site Network Infrastructure

The Base Site network is situated in La Serena, Chile, at the AURA compound. Major network-related functions performed at the Base are: transmission of raw science images and meta-data to the Data Facility; ingest of images and meta-data in the Data Backbone; transfer of control information between OCS, TCS, CCS, and DMS. The raw image data bifurcate in the Base Center computer room for storage in the local Base Center and for forwarding on to the Data Facility.

The bandwidth to the wide area network is a permanent private 140 - 200 Gbps, on which Rubin Observatory transfers approximately 20 TB per night to the Data Facility.

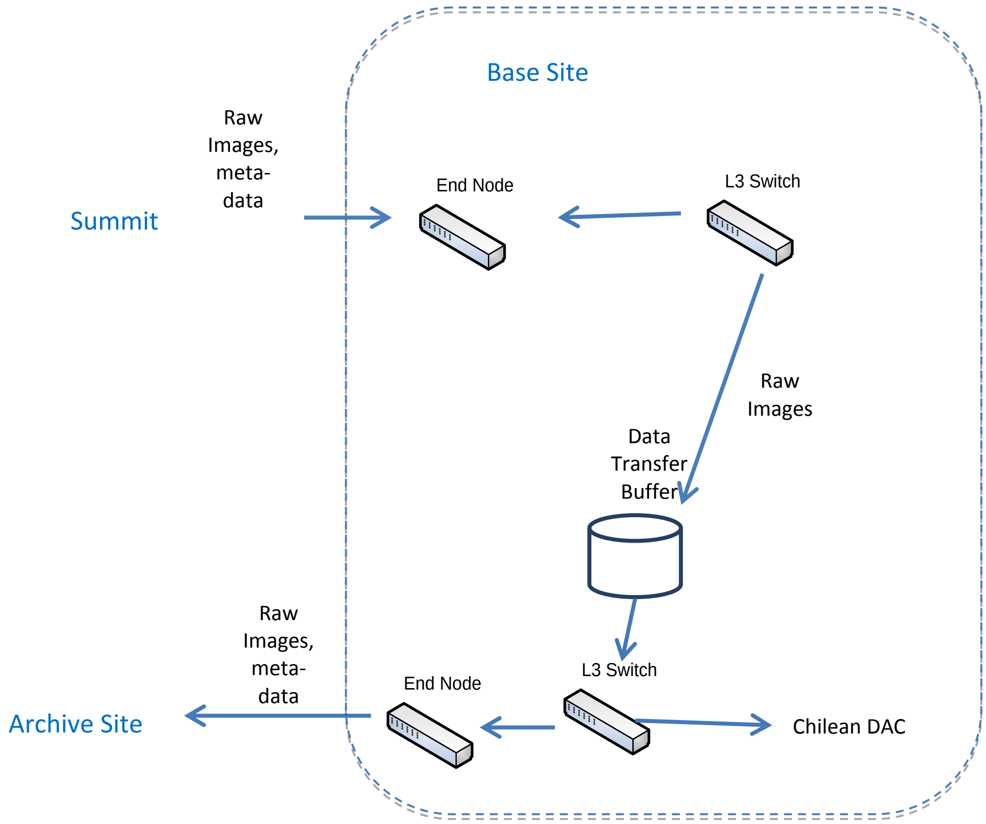


Figure 11 Base Facility Infrastructure Diagram

In Figure 11 the summit connects at the top left of the diagram. An optical image is received from the Camera DAQ divided by raft into 22x10Gbps channels and is converted to 40x10GigE circuits and sent to 10GigE NICs in a rack of replicator servers where they are processed and forwarded to the Data Facility and Chilean DAC for storage.

As described in LSE-299 Base Facility Data Center Design Requirements, each rack houses two top of the rack network access fabric extenders for reliability and redudancy. Each rack is installed with PDUs with individual outlet V/A monitoring capabilities and network switchable power outlets.

It is specified that the only near real-time critical traffic is that of the science image data from the camera to the forwarders in La Serena. All other IP functions can be satisfied by regular switch and routing components at GigE or 10GigE bandwidths.

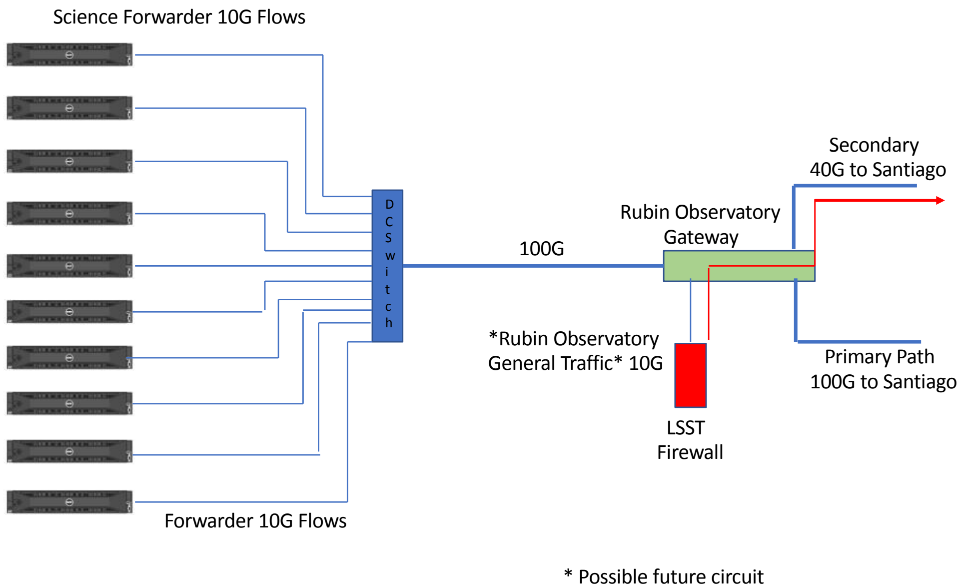


Figure 12 Base Site Science Data Network Infrastructure

Figure 12 illustrates the path of the Science Forwarders egressing to the Data Facility. At some future point the Rubin Observatory general internet traffic may traverse these private paths instead of the AURA network.

6 Chile and International Network

Since 2003, well over USD $40 million (USD $12 million NSF) has been invested in U.S. – Latin America connectivity. There is an existing 100 Gbps ring in place around South America, driven by a wide range of academic and research needs including high energy physics (HEP) and the Large Hadron Collider (LHC) project. Brazil (RNP) has invested internally for 100 Gbps between Rio de Janeiro, Brasilia and São Paulo, and has contributed to the 100Gbps ring.

As shown in Figure 13, this forms a ring that is provided via submarine and terrestrial networks owned by a broad consortium of network operators (including GVT, AmLight/FIU, Silica, Level3, Telecom Italia, and others) seamlessly available to Rubin Observatory through a 15-year IRU and operations agreement with AmLight/FIU. As noted, all the links shown in Figure 13 will be deployed by the end of commissioning, except the 100 G spectrum bandwidth between Boca Raton and Atlanta, which will be deployed early in Operations.

This primary network infrastructure is path diverse from Santiago deployed on either side of South America. The primary Eastern side is spectrum bandwidth on a new fiber connection via Santiago to Sao Paolo to Boca Raton to Atlanta to the Data Facility via ESnet. The secondary Eastern side is via Santiago to Sao Paolo to Miami to Jacksonville to the Data Facility via internet2. On the West coast, Santiago to Miami is currently 100 Gbps with a link from Panama across the Caribbean to Miami.

Rubin Observatory-available guaranteed bandwidth is 100 Gbps on the spectrum link, and a minimum of 40 Gbps in the other links. Provisions for bursting to available bandwidth (up to 200 Gbps) are provided. These bandwidths presume continued investment in these networks by the United States, Chile, and Brazil.

Two providers, REUNA, and AmLight, have been employed to reach the U.S. from La Serena and the entire circuit will utilize DWDM technology. An executed Memorandum of Agreement (MOA) also commits Brazil's RNP REN to support extended Rubin Observatory fiber connections (spectrum) from Sao Paolo to Santiago.

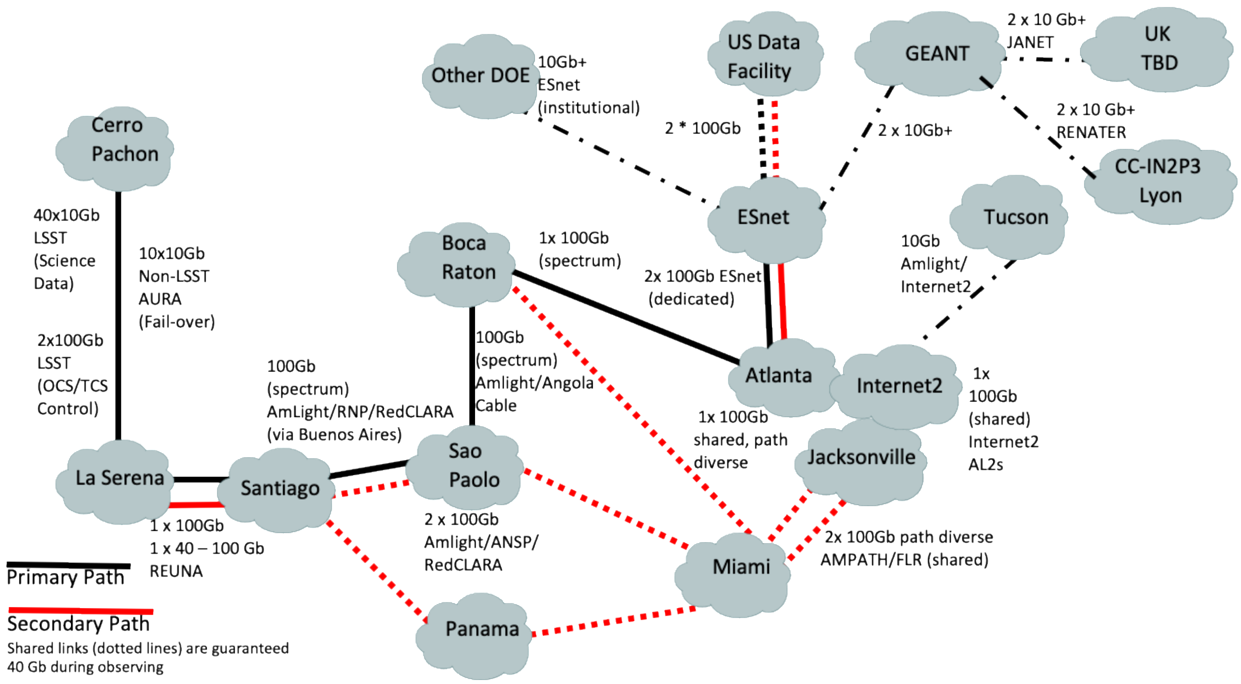


Figure 13 Chile and International Network

6.1 Chile National Circuit

Within Chile, REUNA (Chilean National Research and Education Networks) provides bandwidth from La Serena to Santiago. This link, that provides the last mile circuit from Santiago, transfers Rubin Observatory data between the Base and Archive Sites, and will also be used extensively by the Chilean and Latin American communities to access the Chilean Data Access Center. The Latin American traffic will interact with Chile over CLARA, a South American network.

In 2009 REUNA formed a consortium with ESO named EVALSO and at that time had a 10 Gbps lambda from Santiago through La Serena to Antofagasta. Rubin Observatory provided REUNA with up front funding during the construction phase in order for them to upgrade their network and in return received heavily reduced annual service costs during the LSST period.

This provides Rubin Observatory with bandwidth for the transfer of all Rubin Observatory data between La Serena and the Data Facility.

AURA has had a working relationship with REUNA since 2005 in which it is an active member. Figure 14 below illustrates that there is a dark fiber pair (blue line) in 2017 operated by REUNA with an IEU of 18 years from La Serena to Santiago which is inside a bundle of 96 fibers owned and utilized by the three major Telecoms in Chile. This fiber cable is new infrastructure and an integral part of services provided by the Telecoms to the North of Chile so in the event of failure or cut it will be repaired in the shortest possible time, i.e. 4-6 hours, by the Telecom. There is also a redundant path to Santiago over an existing cable (red line) which initially provisioned 4Gb and will increase to 40 - 100Gb in time for Rubin Observatory commissioning. This is a service provider and not fiber. The bandwidth is available for Rubin Observatory traffic.

Figure 14 La Serena to Santiago link

Figure 15 shows that on the dark fiber pair between La Serena and Santiago REUNA provisioned two OTU-4 optical paths, one (100Gb) exclusively for use by Rubin Observatory and the second for REUNA REN which extends to the north of Chile. Additionally, AURA has reserved the rights for a further 9 Lambdas within the dark fiber for future use. The end equipment in La Serena resides in the Rubin Observatory Data Center and is operated by REUNA.

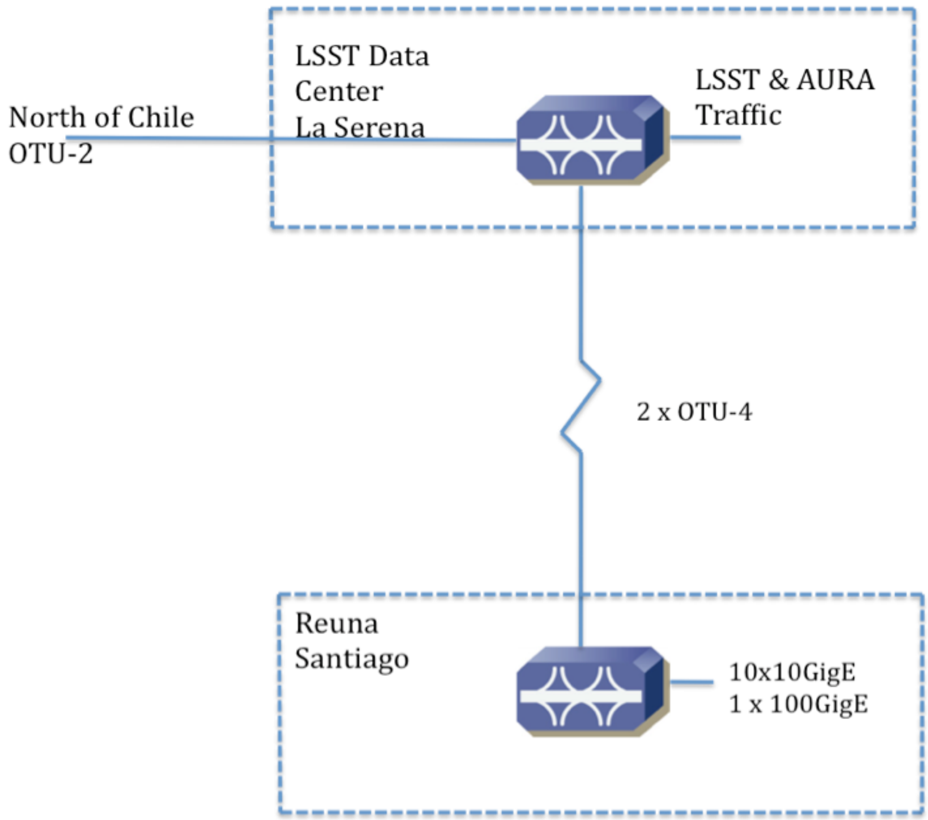


Figure 15 La Serena to Santiago Link

As a means of validating the AmLight and REUNA cost estimates, Rubin Observatory approached commercial companies within Chile directly for bandwidth needs. The three operating telecommunications companies within Chile were approached to offer solutions for this capacity. Two offered viable technical solutions with an average monthly recurring cost of USD $50,000 with availability in 2012. Commercial letters of offer were made, insuring the creditability of the solution. The resultant costs quoted were unsustainable given the funding level, with no opportunity for access to dark fiber and the minimum bandwidth requirements of Rubin Observatory.

6.2 International Circuit

Current carrier technology is typically capable of 100 Gbps on a single Lambda but multiple 10 Gbps per Lambda were the initial implementation on some of these links during Rubin Observatory construction. These are now upgraded to 100 Gbps.

Dedicated bandwidth can be acquired via a services lease, but that means that the carrier decides how to provision the service, rates can change from year to year, and this is typically not the most cost-effective option. An alternative is to acquire Indefeasible Right to Use (IRU) on the lambda or on the fiber. In the former case, the dedicated bandwidth is defined by whatever equipment the carrier is operating at the time. The latter permits the acquirer to put end equipment at any bandwidth on the fiber they desire up to the limits of technology, and increase it at their own discretion.

It is expected that these carrier-provided long-haul links may possibly be upgraded to 400 Gbps Lambdas before Rubin Observatory operations ends. Rubin Observatory's planned investments in long-haul networking during Construction, along with continued investments by AmLight's South American networking partners in Chile and Brazil, as well as potential involvement by commercial entities make this a possibility.

The current design thus specifies 100 Gbps as the baseline dedicated available bandwidth on these links. Further, it is our practice to negotiate IRUs on the fiber wherever possible, so that we are not tied to the carrier's equipment upgrade plans. Rubin Observatory actively participates in the South American Astronomy Coordination Committee (SAACC) that leverages multi-observatory network bandwidth requirements to achieve the best possible terms with the carriers.

Negotiations have taken place such we have letters of intent from all the providers that a ring of 100 Gbps guaranteed bandwidth will be available for use by Rubin Observatory at the end of the construction stage. This will also enable bursting to use available bandwidth (up to 200 Gbps when required, for catch-up. AmLight (FIU) and Global Crossing have both quoted a 15-year IRU giving MTBF greater than 98%. This funding has been paid during the construction period to allay operational costs.

The link from Santiago to São Paulo utilizes a leased carrier line managed by AmLight (as a primary path until the RNP Santiago – Sao Paolo spectrum link). Rubin Observatory contracted an IRU for a 100 Gbps Ethernet circuit that has reliability better than 98%.

The São Paulo to Miami traffic secondary path (managed ring) shares the AmLight network with access to 2 x 40 Gbps, provided by an NSF and Brazilian collaboration. However, Rubin Observatory may “burst” or increase the bandwidth to 2 x 100 Gbps on a capacity available basis for catch-up, and negotiations with the have been conducted in order to secure this option.

7 U.S. National Network

The U.S. national network infrastructure also has two “sides” relative to the path from Atlanta to the Data Facility, as shown in Figure 16. One side begins in Boca Raton at the termination of the international Atlantic side network connection from the Base Site and encompasses the link from Boca Raton on to the Archive Site via the east coast of Florida. The other side begins at AmLight in Miami with the termination of the international network connection from the Base Site and encompasses links from AmLight on the east and west coasts of Florida to the Data Facility.   Each side of the U.S. national network currently supports 40 Gbps with burst capability of up to 100 Gbps, and each side will support 100 Gbps by the end of commissioning.

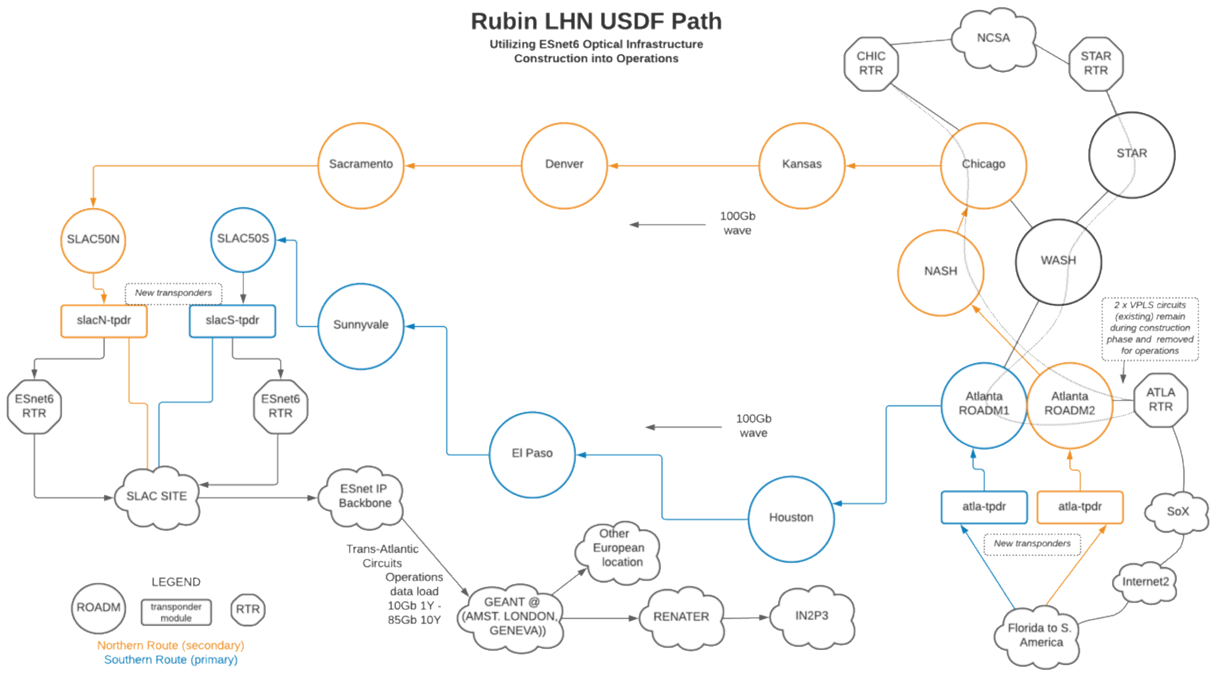


Figure 16 Rubin LHN USD Path

These pan Florida links then traverse north to the ESnet point of presence in Atlanta, and the AL2S port in Jacksonville as a secondary path. The latter is supported through existing FIU-FLR contracts with Internet2.

A new ESnet partnership that is distinct and specifically for Rubin Observatory was executed in FY19. Under this partnership, ESnet will support provisioning and operating two diversely routed 100G Ethernet circuits between the A location, an ESnet POP in Atlanta, Georgia and the Z location, an ESnet POP at SLAC. These 100G waves are built on the ESnet6 optical plant and traverse many different intermediate optical nodes and amplifiers between the A and Z locations. Due to advances in transponder technology, no repeating will be necessary.

ESnet also provides general IP routed peering connections to the Rubin Observatory in the Rubin Observatory Level (3) POPs in Atlanta, El Paso, and Sunnyvale to facilitate collaboration with SLAC, BNL and other DOE laboratories.

ESnet does not usually support transiting traffic between peering connections, but an exception was made for this specific use case for Rubin Observatory science traffic between the Rubin Observatory HQ location and the Data Facility, as well as between the Rubin Observatory HQ and the Rubin Observatory peerings facing the telescope in South America.

ESnet does not currently support a spectrum service but may do so at a future date. As ESnet6 implementation progresses, ESnet will evaluate this possibility and keep Rubin Observatory informed in the event this service is feasible. In the absence of lambda service, multiple 100G transport will be used. It is anticipated over the lifetime of Rubin Observatory the transition to lambdas will be supported.

ESnet will also support connectivity to the Data Facility using higher level protocols should that become advantageous to the Rubin project at some point in the future. Whereas the wave and lambda services are strictly point to point, connectivity built on MPLS would allow more flexibility as to the A and Z endpoints as well as allow additional endpoints should that need arise. 100G service levels would be maintained through QoS policies on the virtual circuits.

8 Trans-Atlantic Network

ESnet is a member of the Advanced North Atlantic Consortium (ANA) and provides 4x100G connectivity between the United States and Europe over four separate and diverse Trans-Atlantic cables. In Eurpoe, ESnet utilizes GEANT's optical ring to provide a 200G ethernet ring between the European ESnet routers. Each ESnet router in Europe peers with GEANT routers in different locations. GEANT connects to IN2P3 via RENATER in France and to the UK processing locations via JANET in the United Kingdom, as shown in Figure 17. These segments are used to transfer raw data from the USDF to the French and UK processing sites, and transfer processed Release data back to the USDF.

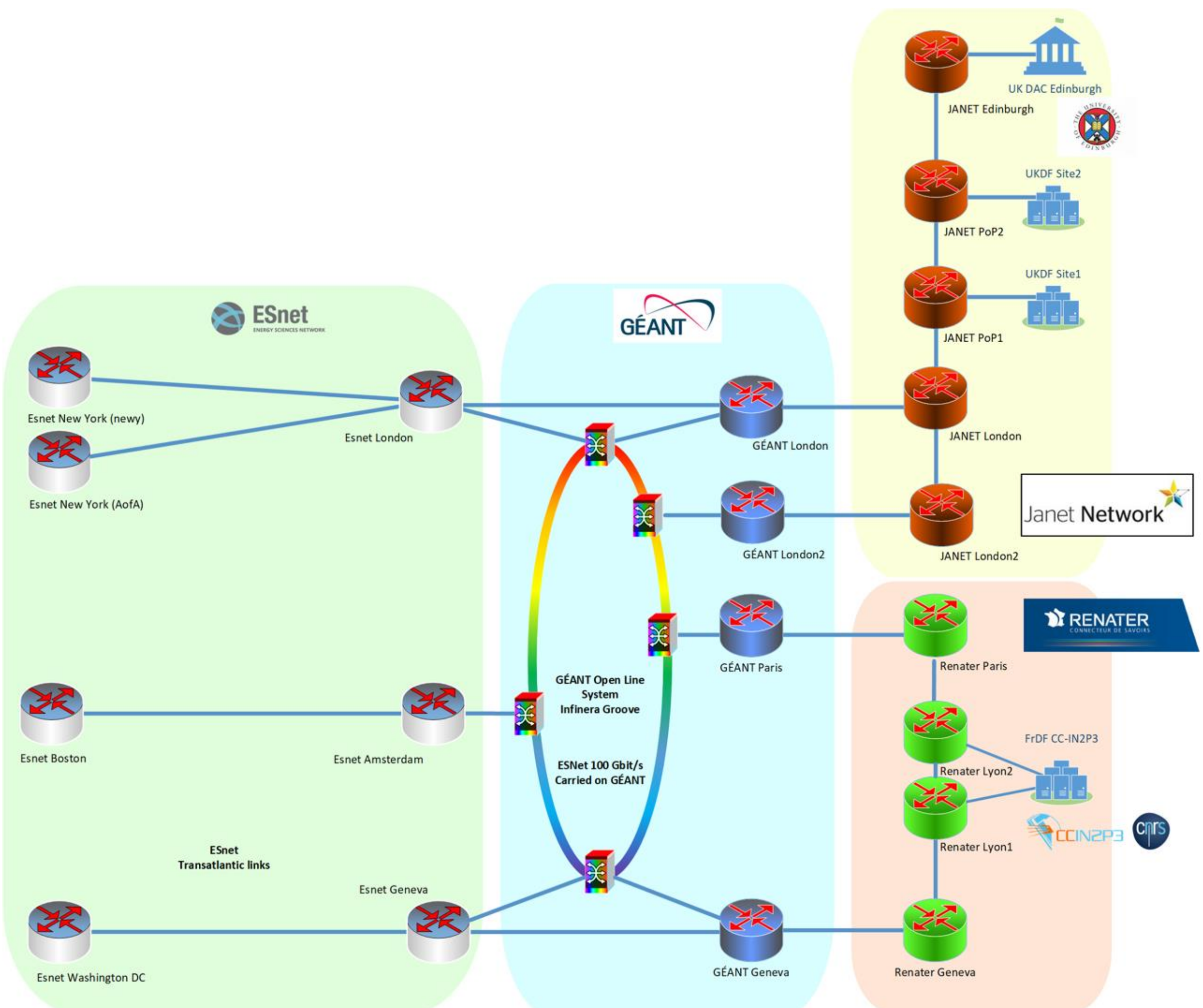


Figure 17 Rubin LHN over Trans-Atlantic Networks

9 Data Access Center Network Connectivity

There are two project-funded DACs, one is in the Base Site and the second is in the Archive Site. The network connectivity between the US national components of the Rubin Observatory, specifically between the Data Facility and the US Data Access Center (within the Archive Site) and between the US Data Access Center and the User Communities are accomplished by the hosting institutions REN and commodity connections.

The Data Facility will transfer large data sets to both Data Access Centers (DAC). Each DAC will have the two most recent data release catalogs, including the Level 1 catalog. Each DAC will have access to the full set of the raw images co-located at that site stored on tape, plus the current sky template and a 30-day cache of the most recent calibrated images stored on disk.

The Data Access Center provides data access and processing services on demand to the various Rubin Observatory User Communities. Note that users do not access the Data Facility or Base Center. Instead, they access all Rubin Observatory data sets via the Data Access Center. The primary data path between the DAC and the institution-based User Community will be across 10 – 100 Gbps institutional links to one (or more) of the Layer 3 based Research and Education networks such as the Internet2 network or the DoE's ESnet network. Each DAC connection will be tuned to support a minimum sustained data transfer rate of 1 Gbps with supporting higher sustained data transfer rates encouraged. User Communities not connected to one of the R&E backbones will access the DAC using the Commodity Internet.

The DAC will be able to support the dynamic allocation of Layer 2 infrastructure utilizing ESnet's OSCARS service combined with Internet2's AL2S Network DCN service, in order to support large data transfers to an end user site. These services allow the allocation of point-to-point VLANS with non-dedicated or dedicated bandwidth for short durations (hours or days).

10 Future Variables

The above networking design assumes the ability to leverage, including cost effective pricing, existing Research and Education Networks. Changes in these networks are expected over the years between now and Rubin Observatory operations, and should be taken into consideration when selecting national and regional network providers and technology. Below is a summary of expected changes over the next few years.

The network connectivity from Chile to the U.S. is being designed to maximize the committed and available bandwidth by collaborating with the AmLight (America's Lightpaths) consortium. The AmLight-ExP consortium (evolved from AmLight and WHREN: Western Hemisphere Research and Education Networks) has been seed funded by the U.S. National Science Foundation's Office of Cyber-Infrastructure (OCI-0441095 and OCI-0963053). This $0.75M annual network funding has historically been through cooperative agreements awarded to the Florida International University (FIU). Significant additional annual funding has been provided by other AmLight consortium members, including funding from ANSP (the Academic Network of Sao Paulo) of $3.8M and funding from RNP (the Federal Academic Network of Brazil) of $3.7M.

There is no reason to believe that the AmLight consortium or its successor (as AmLight is the successor of WHREN) will not continue to operate through the period of operations of Rubin Observatory. However, the NSF cooperative agreements are re-competed every five years, without an OCI commitment to the program in perpetuity. Similarly, the ANSP support is renewed every five years, and the RNP support is renewed every three years.

As such the Rubin Observatory has entered into a contract with FIU/AmLight to provide a minimum of 40 Gbps path diverse capacity from Santiago to Atlanta. This level is contractually guaranteed irrespective of the evolving participation of other AmLight members. It is currently the case that through the continued consortium participation of ANSP and RNP, a bandwidth of 100 Gbps will be available for Rubin Observatory operational use.

Rubin Observatory Operations partner Brazil (through ANSP and RNP) could commit to continue AmLight funding for the benefit of Rubin Observatory through the operational period. If Brazil were to do this, the contract for bandwidth would be able to guarantee protected 100 Gbps. The difference between 40 Gbps and 100 Gbps is significant, but not existential, to the science operations of Rubin Observatory. It improves the reliability to transfer the image data in time to process the data for transient alerts at the Data Facility. Of course, with 40 Gbps, Rubin Observatory can also operate as in the original baseline, with transient alert processing still meeting the 60s latency requirement the majority of the time.

In 2011, Rubin Observatory and the French particle physics institute IN2P3 signed a Memorandum of Intent to have IN2P3 provide computing, storage, communications infrastructure and labor to enable half of the annual Data Release processing to be performed in the CC-IN2P3 computing center in Lyon, France. This includes France provisioning a high-speed network connecting CC-IN2P3 and the Data Failicty. There are existing connections between CC-IN2P3 and Chicago via the Advanced North Atlantic (ANA) Consortium and ESnet networks, and these connections provide the foundation for the CC-IN2P3 Data Facility link.

Appendix A: La Serena – Santiago Network Rings

Several institutions have already started initiatives to create path diversity between La

Serena, Antofogasta, and Santiago Chile that Rubin Observatory may leverage, as depicted in this Appendix. Leveraging these rings would further increase the reliability of the Rubin Observatory data transfers and provide additional security against a cut on the traversal of the Andes, which has traditionally been the area most prone to such outages.

There are at least 2 distinct efforts currently underway to establish a ring including La Serena and Santiago, both of them involving Andean crossings and having part of the link in Argentina. There is also a ring that may connect Santiago with Buenos Aires that may be leverage by Rubin Observatory. AmLight is coordinating with these efforts to leverage such a ring for Rubin Observatory use, via the SAACC and direct discussions.

In addition, we are tracking another option using a new submarine fiber between La Serena and Valparaiso that could be a reliable alternative.

ALMA (Paso de Jamas)

Alma Site

This ring contains several segments.

1. ALMA OSF to Calama - New fiber install, currently 10 Gbps, joint funded with Silica, Silica owns and operates, but ALMA has (25 year IRU) on one dark fiber. ALMA (or REUNA) operates end equipment on that dark fiber.

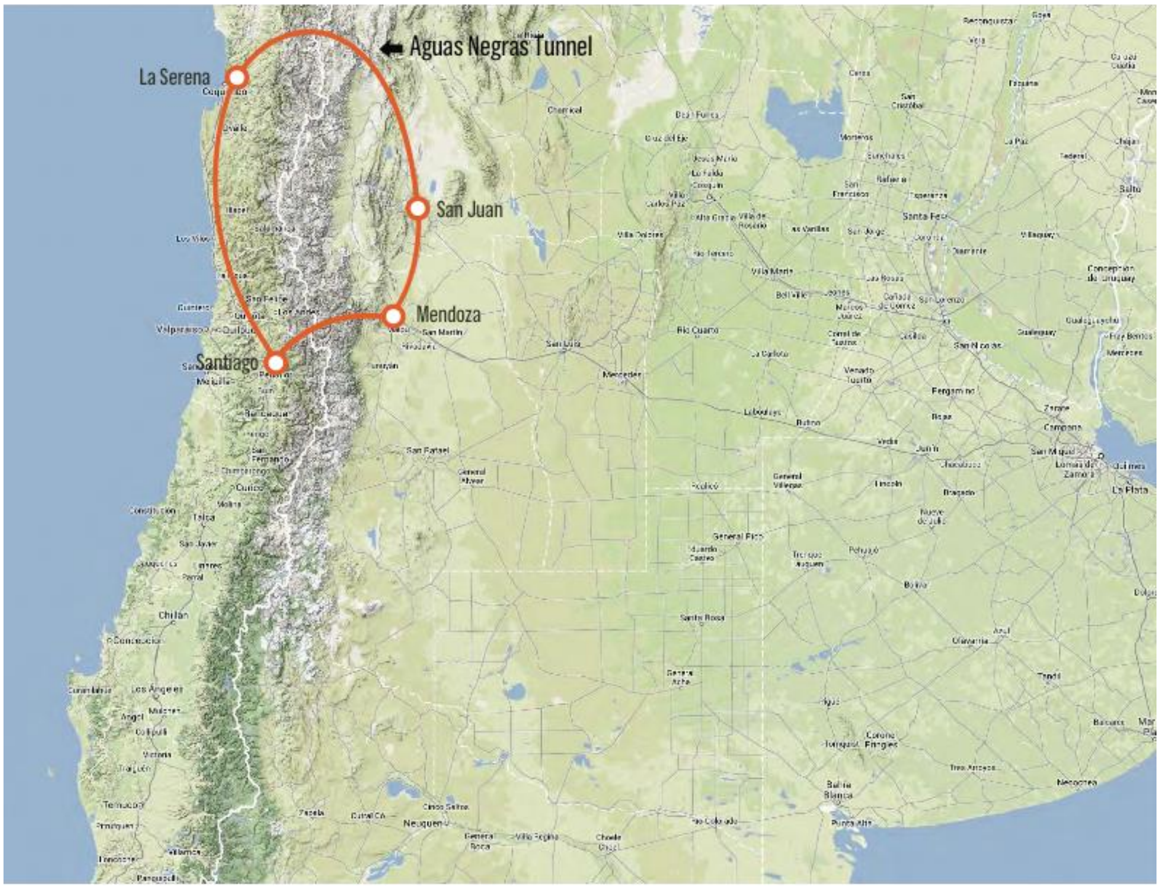
2. Calama to Antofogasta: Telefonica owns fiber and operates. ALMA funded, currently 2.5 Gbps, IRU on the lambda, upgraded if carrier upgrades to 40 Gbps.

3. Antofogasta to Santiago: EVALSO funded. Telefonica owns and operates fiber. Currently 10 Gbps (split into 4* 2.5 Gbps channels). IRU on lambda through 2020, upgraded if carrier upgrades to 40 Gbps. Expect to happen in 2 to 5 years. Equipment last 5 - 7 years, and had been in use for ~2 years before this IRU started. To increase before carrier is ready, have to add equipment or buy another lambda. REUNA currently operates end equipment to split into channels, would need separate investment on this equipment to go to 40 Gbps. ALMA would prefer to change IRU to fiber, rather than lambda, to maximize flexibility for all users. Also discussing with REUNA to upgrade fiber.

4. OSF to Santiago via Argentina (Paso de Jamas): Silica owns fiber. Expected within 5 years. Elastic capacity, 10 year IRU for one lambda, when carrier upgrades will upgrade to 40 Gbps. Currently "gentleman's agreement" to only use as backup, not as another primary path, but inquiry about 100 Gbps was considered not impossible.

Aguas Negras

There is another planned ring connecting La Serena and Santiago going up Elqui Valley and via the planned Aguas Negras tunnel into Argentina. This was presented to us by Florencio Utreras of Alice/CLARA. It requires a planned $800-900 million USD investment, apparently mostly Argentinian, that has to be made to make this tunnel a reality first. Thus the Paso de Jamas is considered more "real" as it doesn't require a tunnel, but instead simply runs along an existing pipeline for most of the way. This path would give also path diversity from both Cerro Pacon to La Serena, making it even more attractive. Rubin Observatory could share installation of fiber on public road, and have an IRU on some number of fibers. Other paying customers on this fiber will likely drive a fast MTTR to meet our specifications. There is also a neutrino lab that might provide funding on this link.



Aguas Negras Tunnel

Chile-Argentina Ring

AmLight is also in discussions with Entel and other entities about completing a third ring, and it is considered likely to be available in Rubin Observatory's operations period. The solid lines are existing and owned by Silica. The ring is partially complete with ENTEL at this time;

meaning that this map shows dark fiber, there is additional leased capacity that makes this work. Odds are very good that the ring will be complete because of commercial interest. There are detailed plans to complete a Malargue to Santiago link.

Chile-Argentina Ring

References

[Freemon2013] Freemon D. M.: Optimizing Throughput on Guaranteed Bandwidth WAN Networks for the Large Synoptic Survey Telescope (Rubin Observatory). CoRR abs/1303.7467 (2013)

[Rubin Observatory14313] Dossa, D., Freemon D.M., Kantor J., Shipping Media Analysis, rev. 2013

[Rubin Observatory11811] Lambert R., Feasibility of Redundant Fiber from Mountain to Base, rev. 2013

[Rubin Observatory11918] Ibarra J., Cox C., Jacque S., Lambert R., Kantor J., Rubin Observatory Network Operations and Management Plan, rev. 2013

[LSE-30] Rubin Observatory System Specifications (OSS)

[LSE-61] Data Management System Requirements (DMSR)

[AURA Chile MOA] Memorandum of Agreement between AURA and Chile

[AURA Rubin Observatory MOA] Memoranda of Agreement between Rubin Observatory and AURA