



Data transfer from ALMA to North America

Adele Plunkett, David Halstead

National Radio Astronomy
Observatory



www.almaobservatory.org



NRAO and ALMA

ALMA is a multinational project with many partners, and three ALMA Regional Centers (ARCs):

- NA: NRAO, Charlottesville, VA, USA
- EU: ESO, Garching (Munich), Germany
- EA: NAOJ, Mitaka (Tokyo), Japan

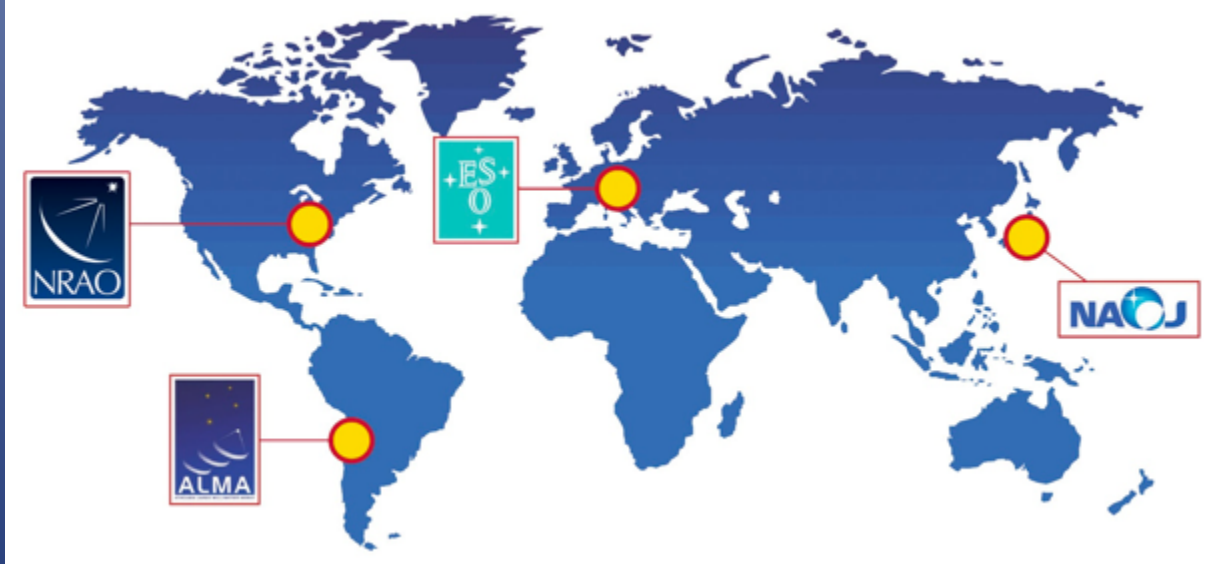


Image: <https://science.nrao.edu/facilities/alma/images/arcs.jpg>

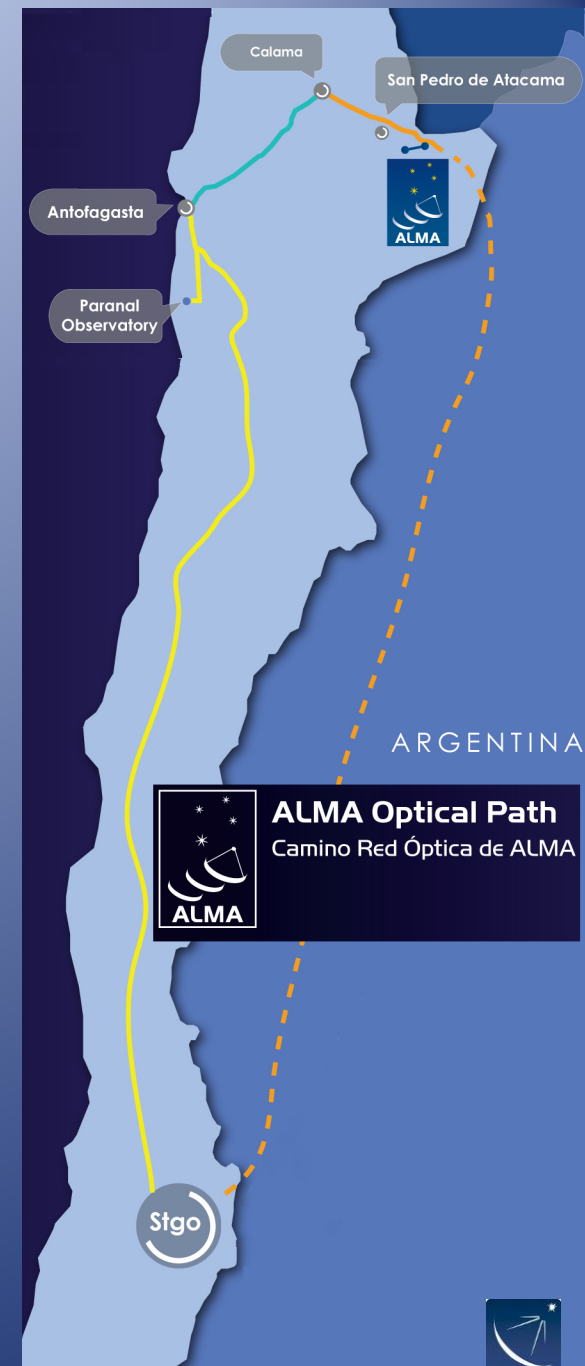
ALMA telescope

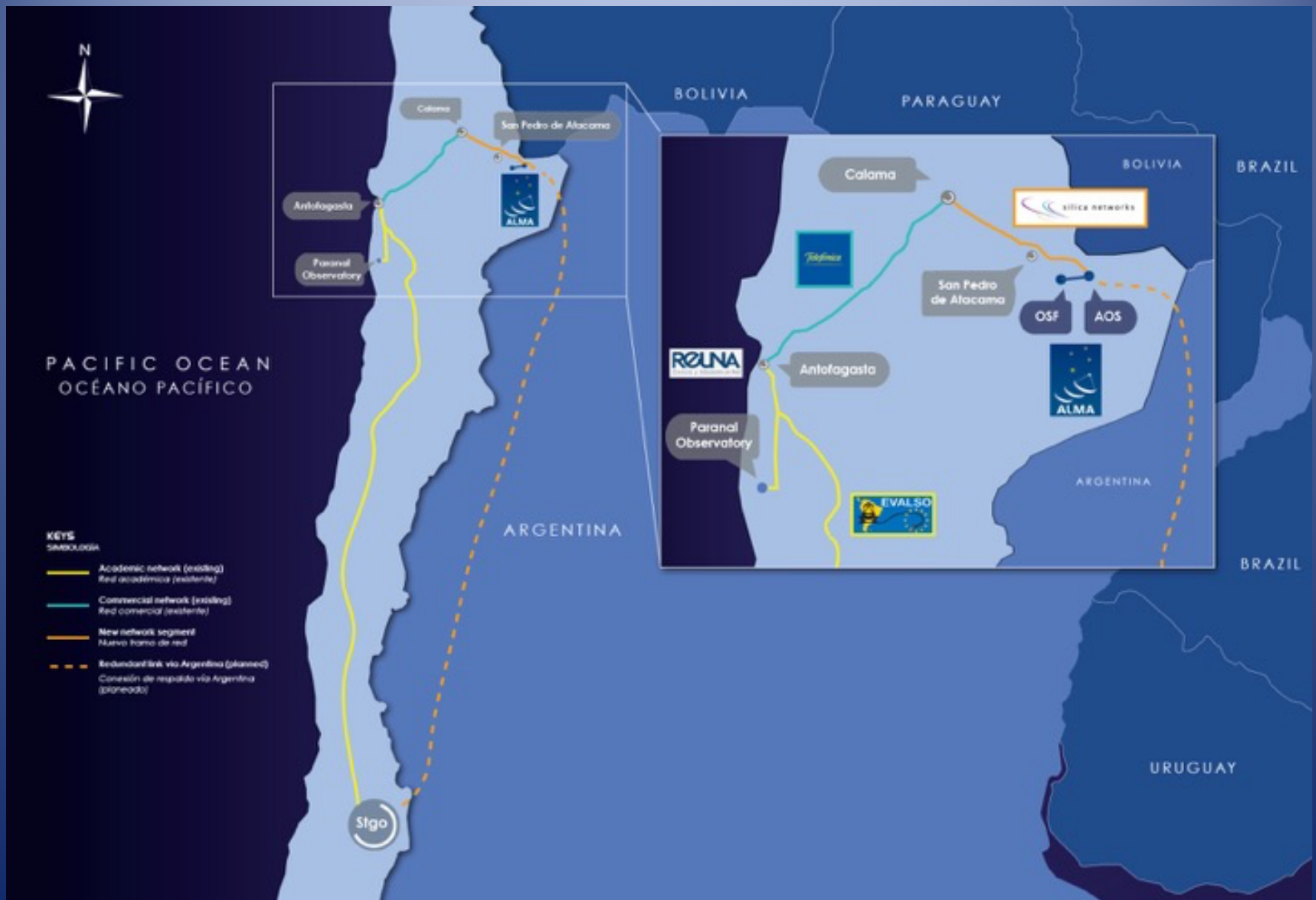
- Largest mm/submm telescope ever built. As an **interferometer**, it combines signals from multiple antennas to form an image.
- All 66 antennas operational at high site (except for maintenance)
- Sub-arrays possible, and generally **3 projects observed at once**:
 - 50 x 12m-antennas (main array)
 - 12 x 7m-antennas
 - 4 x 12m-antennas, observing in “Single Dish” mode (“Total Power”)
- Operated “space mission” style, with pipeline data processing and a science archive at each ARC (and Chile).
- Recent “Control Room Extension” at Santiago Central Office.
- First PI projects released to public from the ARCs January 2013. Annual calls for proposals. **Cycle 7 observations began in October 2019.**
- *ALMA currently shut down since March 22, not observing amidst COVID-19 danger.*

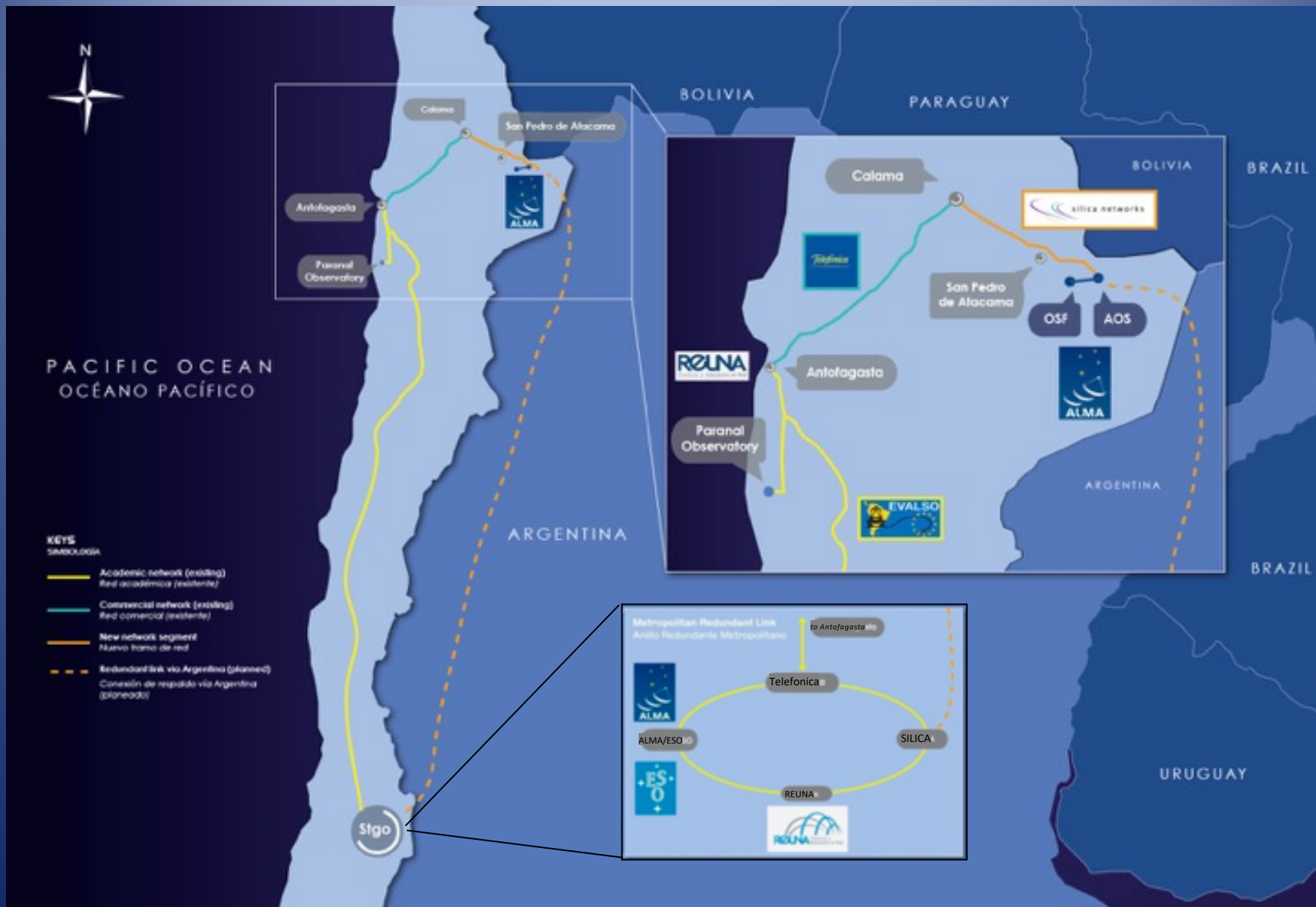
Data transport within Chile

From the telescope

- AOS to Santiago 2.5Gb/s
 - fiber to Calama,
 - commercial fiber Calama to Antofagasta,
 - EVALSO/REUNA from Antofagasta to Santiago
- Redundant fiber loop via Argentina planned







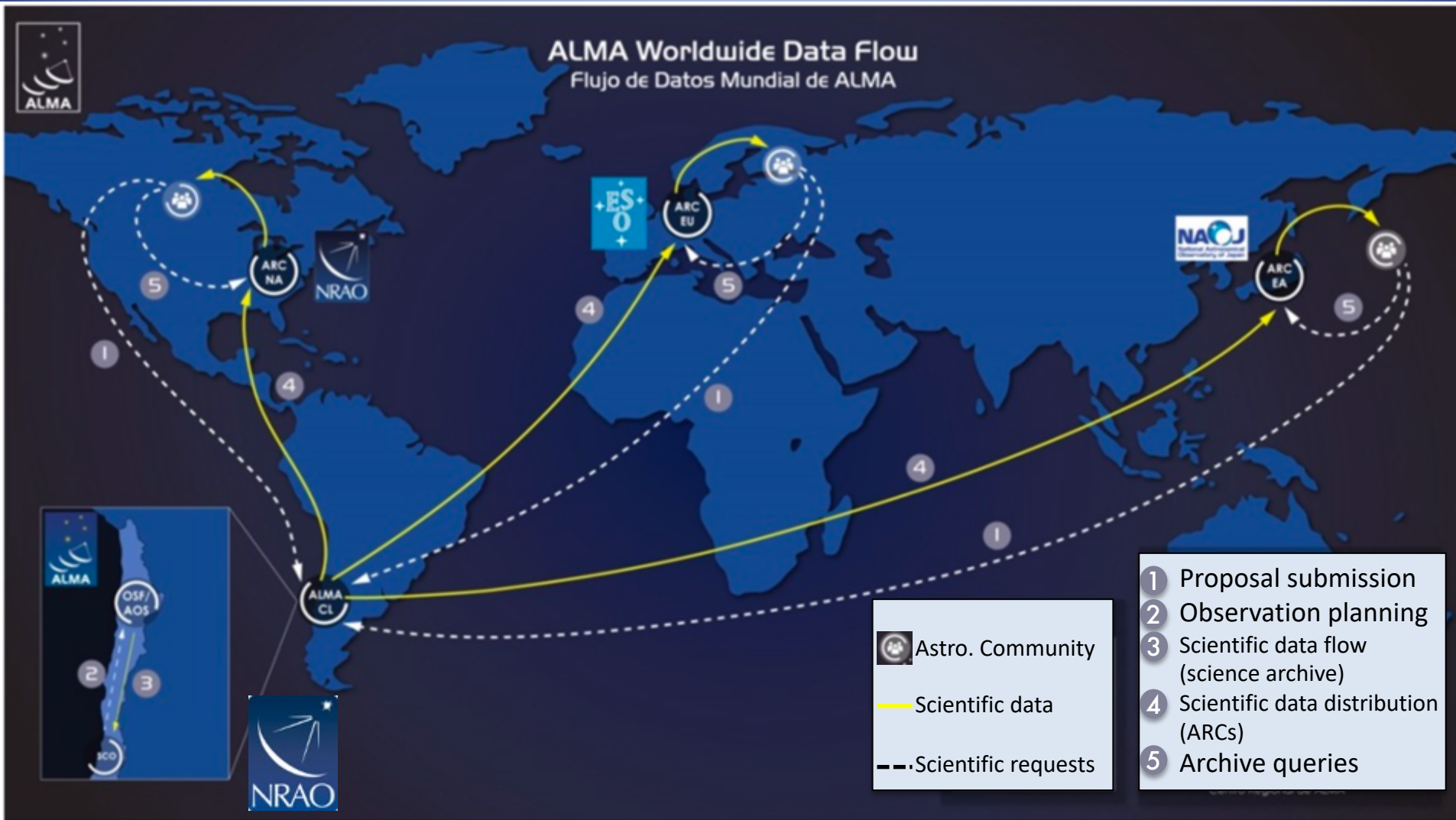
Data transport

Chile to Charlottesville (NAAASC)

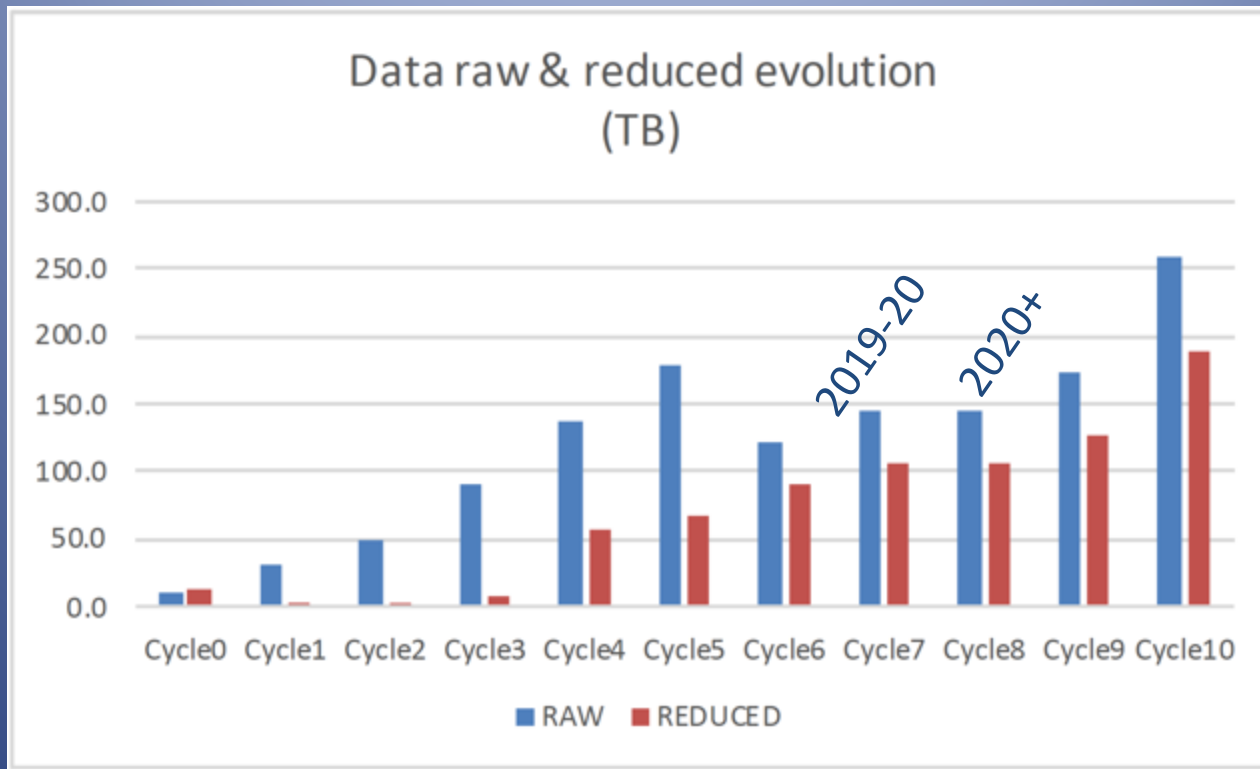
- MOU signed between AUI/REUNA for local link to SCO
- Santiago to ARCs: individual ARC contracts with REUNA and NRENs.
 - NA: Joint AURA-AUI agreement for 1 Gbps committed (burstable to 10 Gb/s capacity) from Santiago to US NREN via Latin America's Nautilus Point of Presence
 - Link from NRAO to Internet2 through UVa is 10Gb/s
- Typical rate obtained during peak data transfer periods is 2-300Mb/s, with bursts up to 600Mb/s
- Currently working on establishing network monitoring, and improving our understanding of how the link performs in typical load conditions (~1TB/day)
- Recent transition to most data processing being done in Santiago. Some manual processing at ARCs.
- The North American ALMA Science Center (NAASC) hosts the ALMA Archive, and computing for NA users.



Data transport (cont'd)

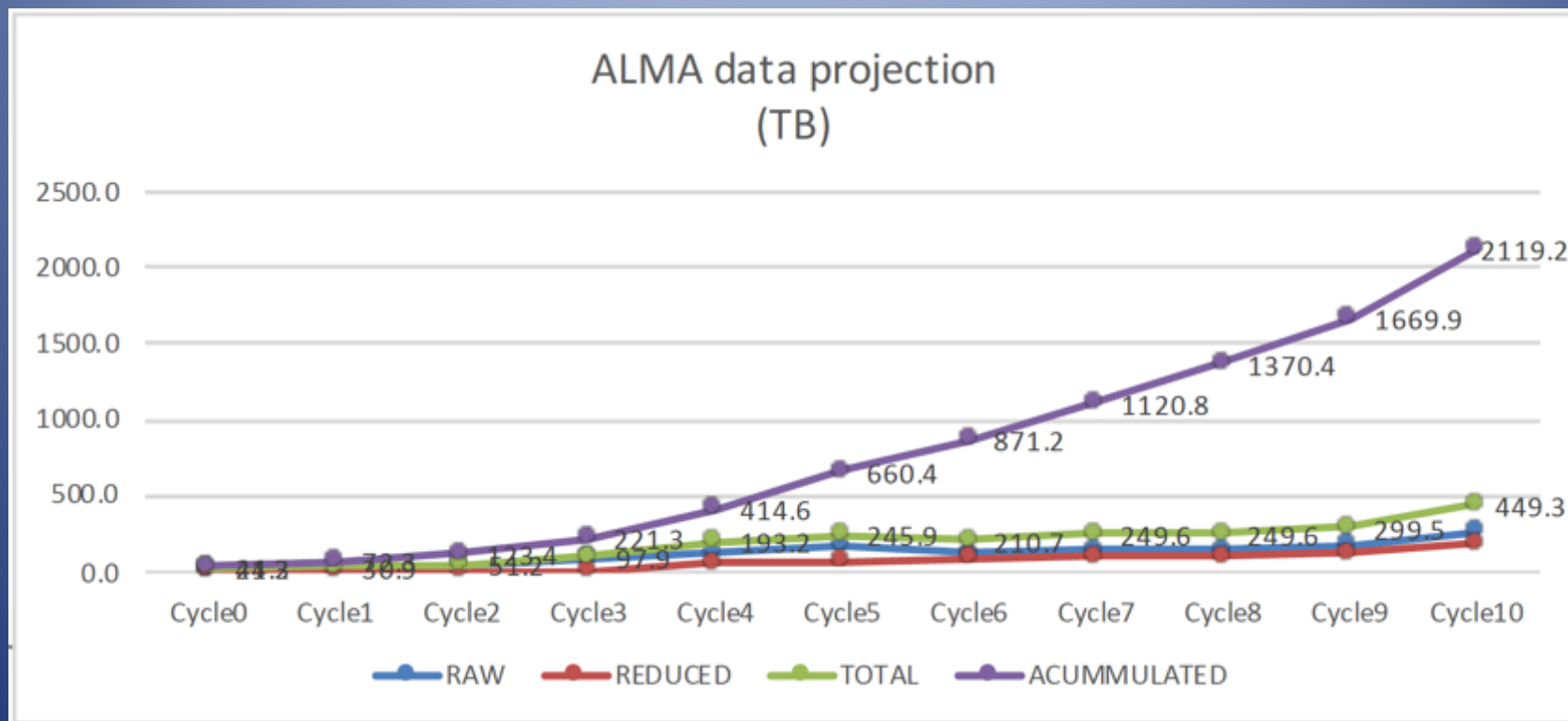


ALMA Science data storage



Source: ALMA data rates and archiving at the NAASC NAASC Memo 110 (v5.1)
ALMA Archive Review, 2019

ALMA Science data projection



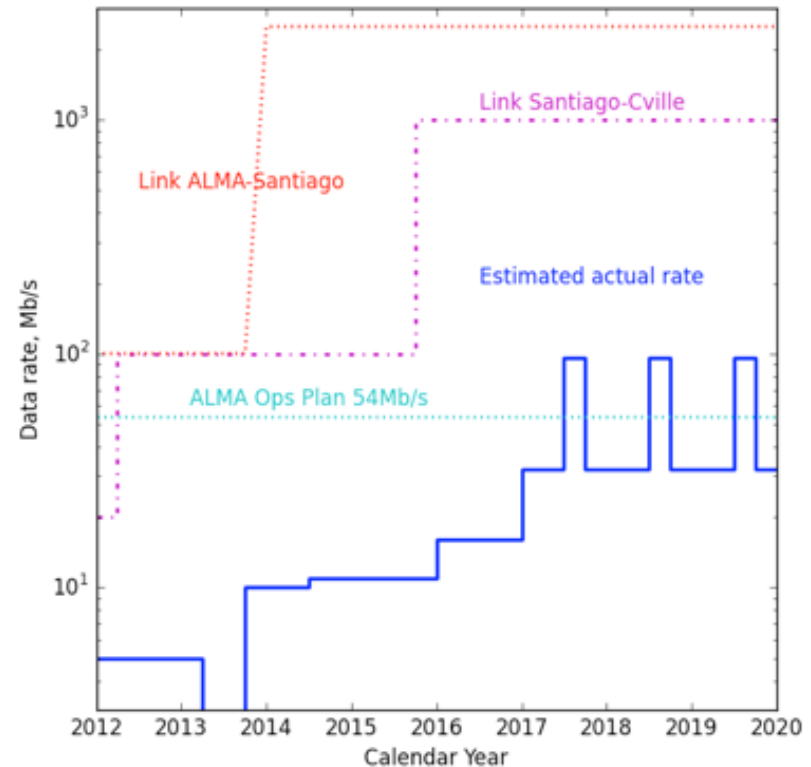
Source: ALMA data rates and archiving at the NAASC NAASC Memo 110 (v5.1)
ALMA Archive Review, 2019

Current data rate, and projections

- Blue (solid) line shows data generation
- Assumes no imposed limit on data rate (cyan dotted line is general Operations Plan rate)
- Data transmission is per ARC
- Does not include emergency data replication (NGAS or Oracle DB)

Note: This plot is an estimate, to be updated

See: NAASC Memo 110, Lacy, Halstead, Plunkett



ALMA science and data rate evolution

Cycle	Dates	Antennas ^[1] (12m 7m)		Array time for science	Data vol. (TB)	Notes
0	Oct 2011-Jan 2013	16		5-10%	20	•“Early science”
1	Aug 2013-Jun 2014	32	7	10%	40	•Inauguration March 2013 (https://www.eso.org/public/news/eso1312/)
2	Jun 2014-Sep 2015	34	10	15%	70	•Some carry-over from Cycle 1
3	Oct 2015-Sep 2016	36	10	25%	140	•Mix of manual and pipeline processing, products ~20% of data vol. •2 data streams (WVR correction)
4	Oct 2016-Sep 2017	40	10	33%	210	•Mix of manual and pipeline processing. •2 data streams
5	Oct 2017-Sep 2018	43	10	45%	275	•Imaging products ~30% of data volume •2 data streams
6	Oct 2018-Sep 2019	43	10	55%	220	•Single data stream •Transition to more pipeline processing at JAO.
7	Oct 2019-	43	10			•Most imaging processing by pipeline at JAO. •Imaging products up to ~50% of data vol.

[1] Minimum number of antennas in 12-meter and 7-meter antenna arrays. In full science, up to **50 x 12-m antennas** and **12 x 7-m antennas** may be included in the arrays; in addition to the “Total Power” array of 4 dedicated 12-m antennas.

Cycle 7 (interrupted)

From: Sean Dougherty

Subject: ALMA Operations shutdown

Date: March 18, 2020



Dear colleagues:

With the current developments in the country related to coronavirus, the CMT has decided to start preparing for a total shutdown of the Observatory over the coming days. This will enable JAO staff to comply with request of the authorities to reduce contact with others to reduce the spread of the virus and to protect yourselves and your families.

The plan is:

- All SCO-based staff should be working from home, unless critically required to be at SCO e.g. an AoD in the SCO Control Room until Thursday evening and a member of the CMT to act as SCO MoD each day. Please coordinate your priorities your work from home with your supervisor and do not hesitate to ask for any technical support you may need to make remote work effective.
- The number of staff and contractors at the OSF is gradually being reduced. Any travel changes, including those for self-arranged travel, will be duly reimbursed. The flight situation is very dynamic, so please be patient and be prepared for changes in flight schedules.
- By Friday March 20th, only those essential staff needed to support the safe shutdown of the observatory will be onsite, led by Stuartt Corder. The focus will be the AOS shutdown and to bring the transporter down to the OSF.
- By Saturday, we expect to complete the AOS shut down. A small power generator will keep the maser powered up in order to protect its stability. All other equipment will be powered down.
- On Sunday, the only people onsite will be those required to complete the IT and Power System shutdown tasks and to lock down all buildings and equipment to ensure all assets are protected. A small generator will provide power to the fuel station and to the water treatment plant. All other systems will be off.
- By Sunday noon the last staff will leave the site.

Future Cycles

- Now running in “Full Science” state, with mean data rates $\sim 100\text{Mb/s}$ during observations
 - “Duty cycle” of observations will slowly increase as testing and maintenance procedures improve.
- Best guess estimate for the next 3 years (including product size mitigation) is around $200\text{-}300\text{TB/yr}$ (raw data and products roughly equal)
- Important to note that data rates vary through the configuration cycle. When long baseline configurations are scheduled the data rate goes up for these reasons:
 - Data sampling needs to be faster to prevent beam smearing at the field edges
 - The data products, which are also mirrored from Santiago, also increase in size, to become larger than the raw data in the largest configurations
 - So far, long baseline campaigns have tended to have low observing efficiencies, however this may change

“Next ALMA Correlator”

- A correlator upgrade is being developed
 - In line with ALMA Development Roadmap [1], and ALMA2030 vision
 - Working group established
 - Ongoing workshop series [2] to refine the specs
- Deployable location physically separate from BLC (potentially OSF)
- Increased bandwidth (~8x more channels)
- Expected data rate increase is about a factor of four, corresponding to a data rate of ~1PB/yr (not all projects will need the extra channels).
- Details to be defined.

[1] <https://www.almaobservatory.org/en/publications/the-alma-development-roadmap/>

[2] <http://www.cvent.com/events/the-alma2030-vision-design-considerations-for-the-next-alma-correlator/event-summary-8c936ac86f8a4e51842644ae58e2d315.aspx>

Summary

- Ramp-up of the ALMA data rate has been slower than anticipated, allowing us to stay ahead of the curve.
- Data flow mostly from JAO to ALMA Regional Centers, with data processing mostly at JAO.
- Still monitoring how the network performs when transferring $\sim 10\text{TB/day}$ in multiple parallel streams.
- Would like to establish a link with 10Gb/s available bandwidth (out of a 100Gb/s pipe) within the next 1-2 years to improve transfer speed to and from Chile for bulk reprocessing, and to help with occasional large data and metadata transports (e.g. a DB export)
- Most new developments (e.g. next generation correlator) on $\sim 5\text{-}10\text{yr}$ timescale can probably be accommodated without increasing the data rate by more than a factor ~ 4 .