NRAO's New Archive and VO

Experiences integrating VO into the AAT/PPI

Stephan Witz

Science Support & Archive Team Lead

Atacama Large Millimeter/submillimeter Array

Karl G. Jansky Very Large Array

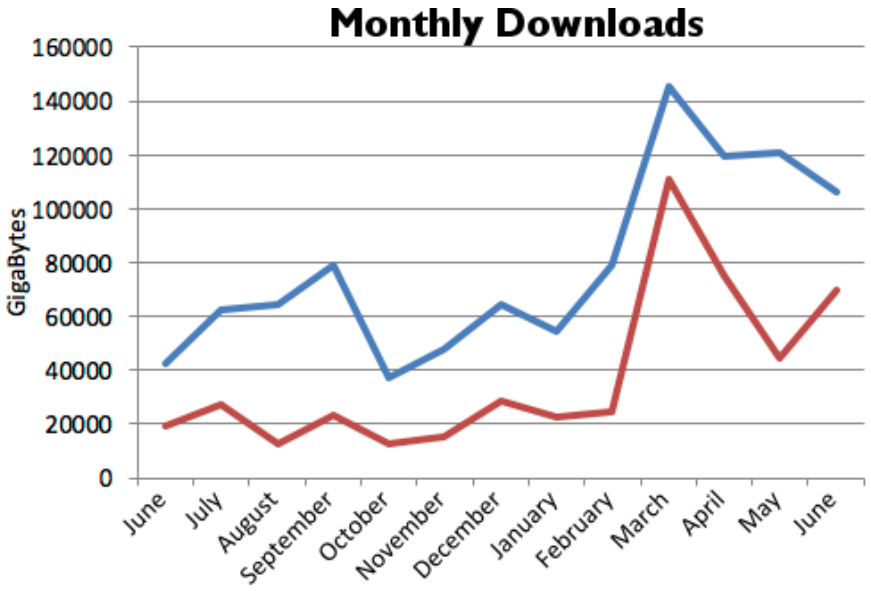
Robert C. Byrd Green Bank Telescope

Very Long Baseline Array

New Archive: Requirements & Roadmap

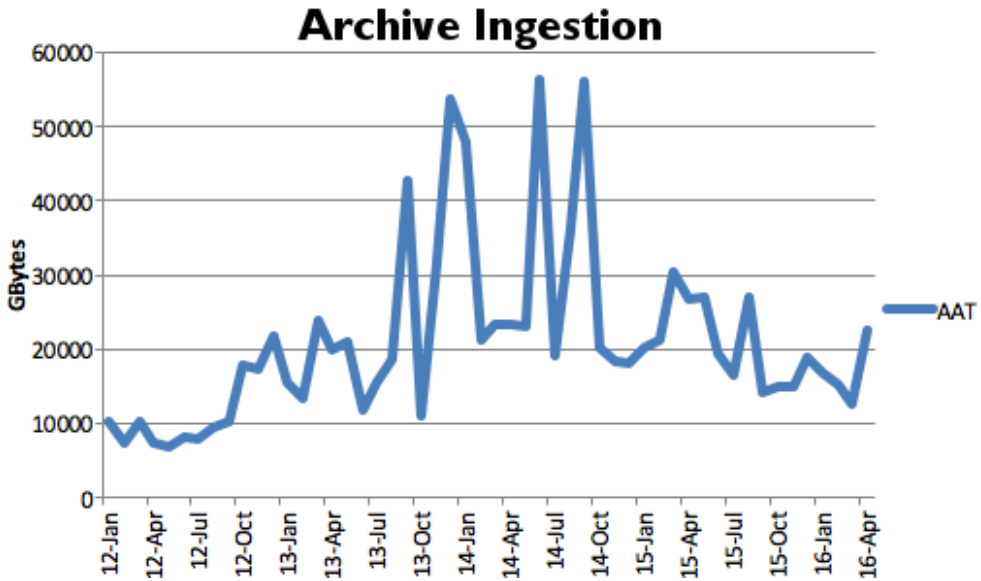
- Requirements for next generation NRAO Archive included:
  - Query ALMA & non-ALMA (legacy VLA, Jansky VLA, VLBA & GBT),
  - Authentication & authorization for proprietary data, ALMA's CAS server for ALMA data, NRAO's CAS server for NRAO data
  - Push the button with no other input, get back a list of everything, ~600k file sets spanning 40 years
  - User-specified reprocessing using CASA and NRAO's cluster
- Phase 1: September 2015, limited proof of concept with all the major components hooked up
- Phase 2: September 2016, Phase 1 + more searching options, new responsive interface, authentication + authorization
- Phase 2.5 & 3.0 delayed by VLA Sky Survey, expected January & March 2017, more processing options, production deployment

Archive Usage & Ingestion

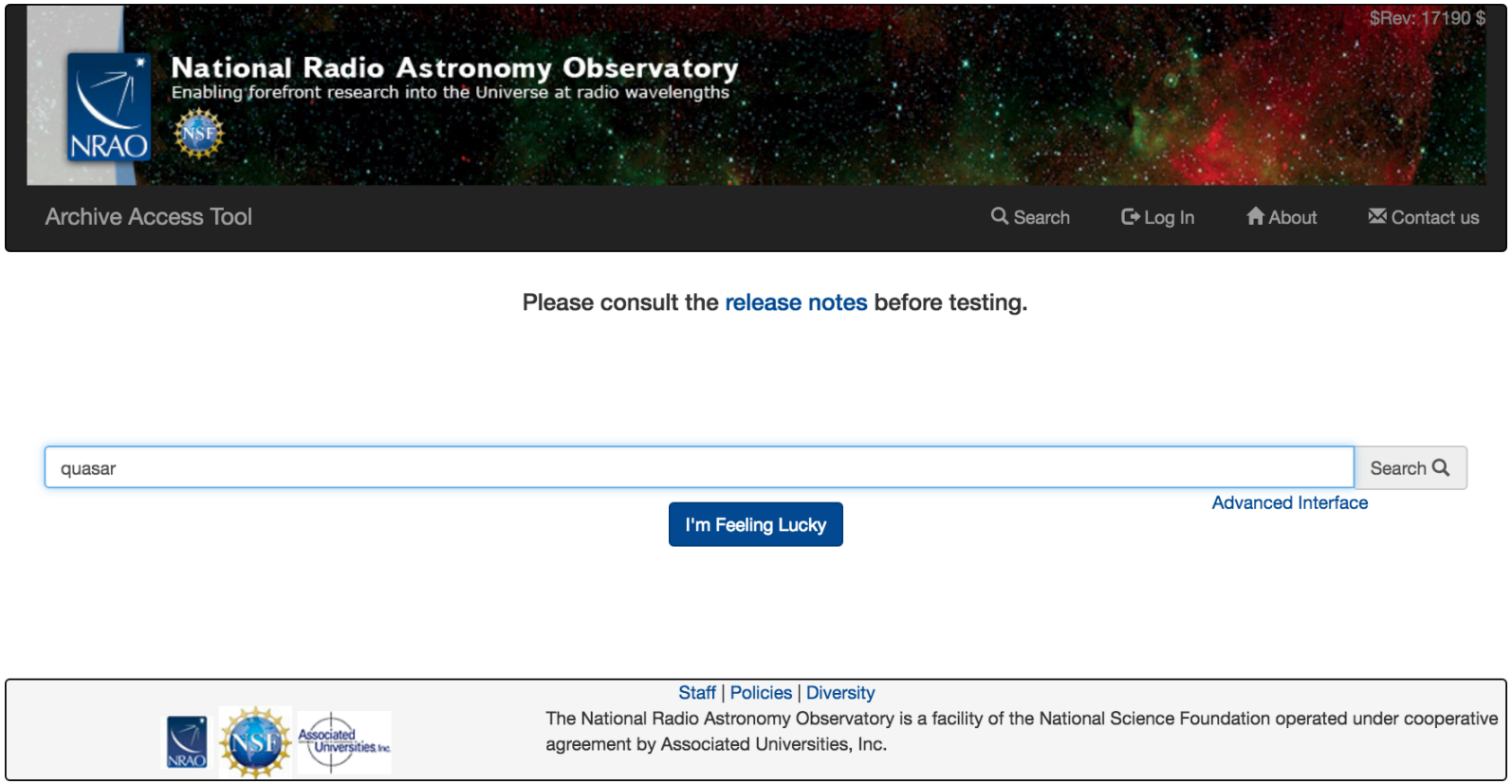


- ~75 users downloading files a day
- ~ 500gb – 2tb a day

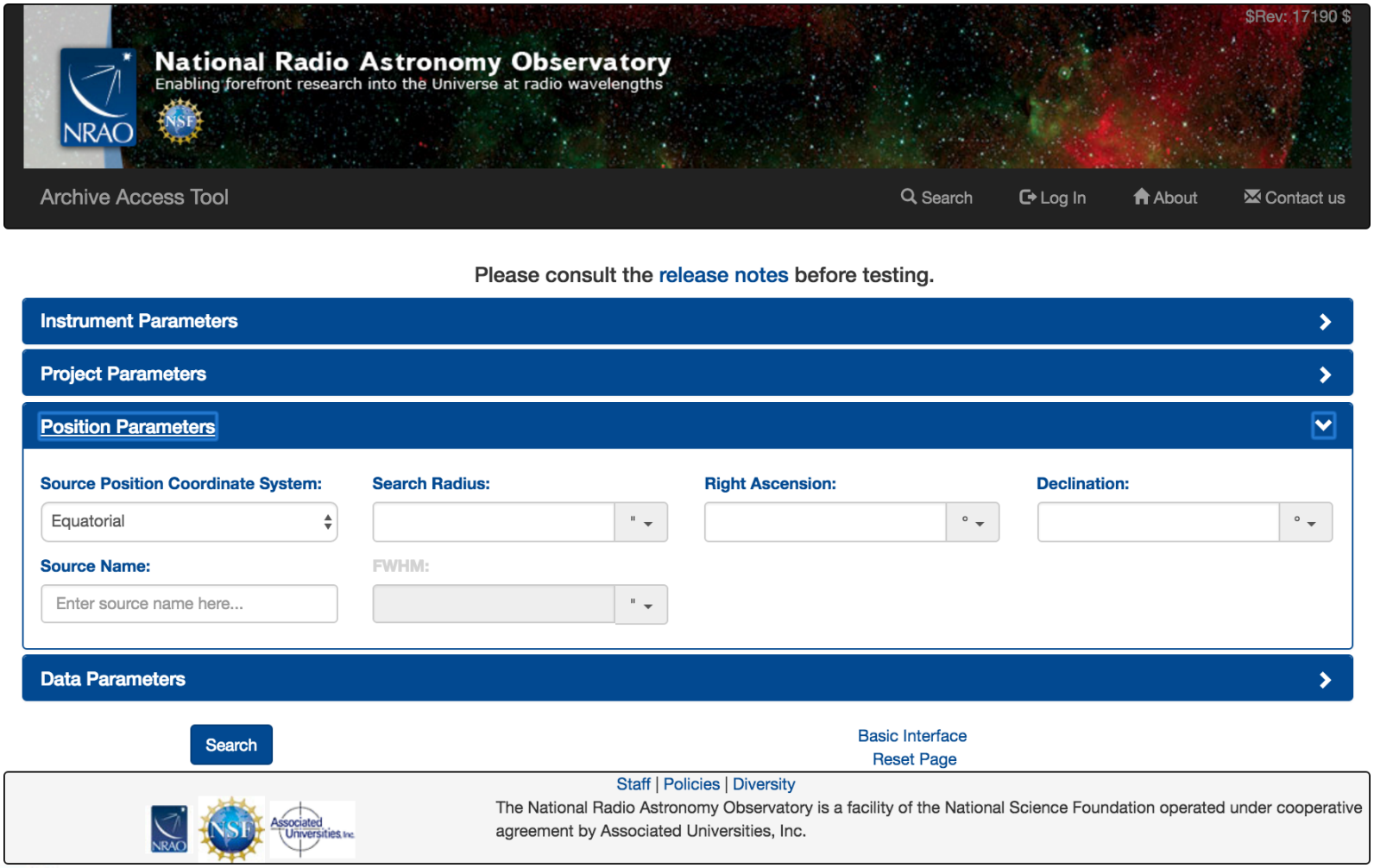
- 96/3/1% VLA/VLBA/GBT



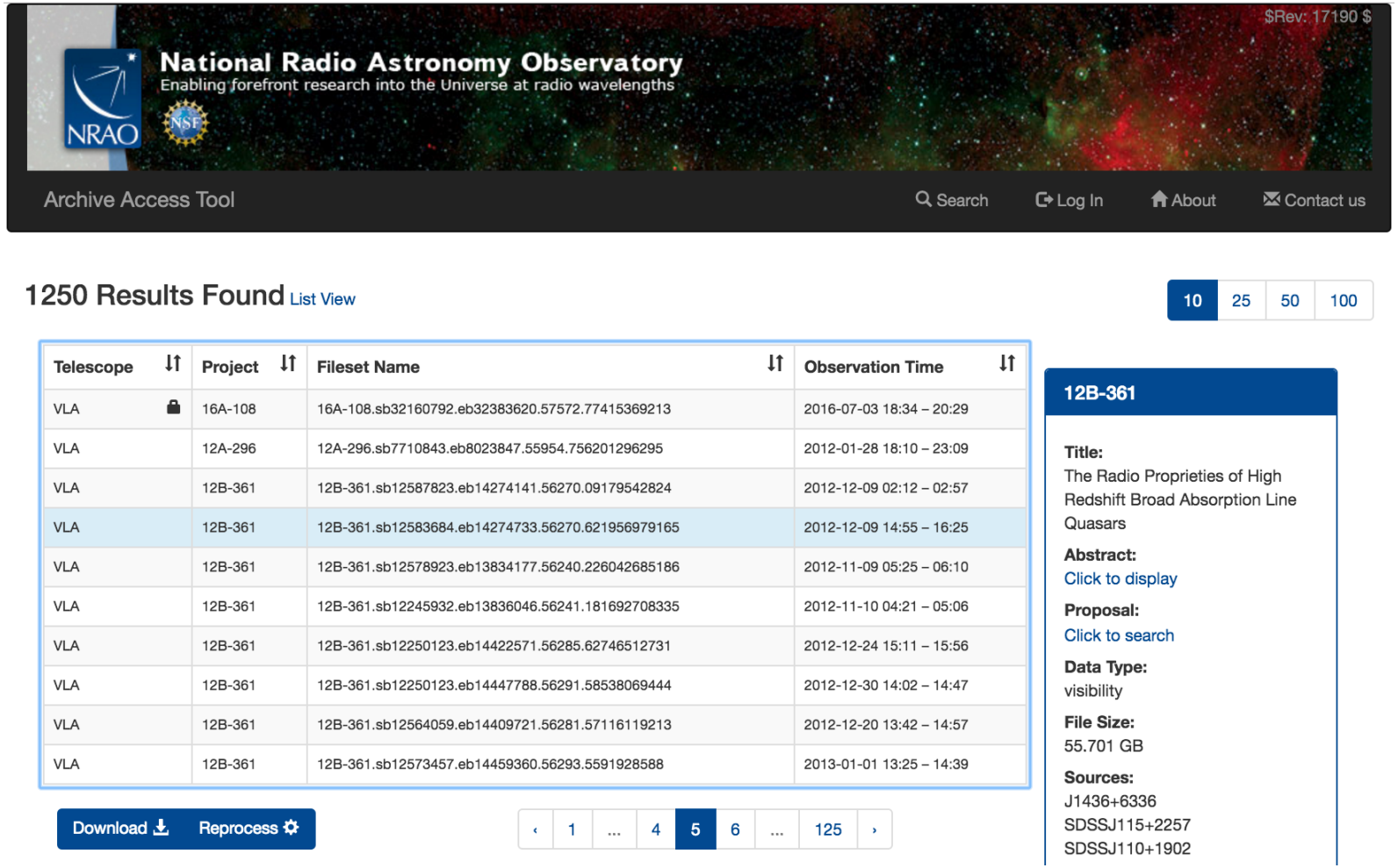
Basic Query Interface



Advanced Query Interface



Results: Table Display



Phase 1: Design

Phase 2: Preliminary Design

Front End

Request Handler

OODT

Cluster Scheduling System

Merge TAP results

NRAO MD DB & obscore view

ALMA MD DB & obscore view

ALMA NGAS

NRAO NGAS

Phase 2: Problems

"*Everybody has a plan until they get punched in the mouth.*" Mike Tyson

- Loss of two key personnel due to retirement, our local contact for VO technologies & chief developer for current archive and VO
- ALMA's DB is Oracle, NRAO's DB is PostgreSQL
- Implemented front end as SPA with dynamic pagination, needed LIMIT and OFFSET, no ADQL OFFSET capability, so different query for each DB (rownum, Oracle, really?)
- Merging VO tables problematic, sorting by relevance with two results sets?
- Struggled with display (part of this likely was team transitioning from Java server code to Javascript SPA & Angular.JS)
- As mentioned, Javascript SPA really wants JSON instead of XML
- We tried SOLR as a way to bridge the gap between the two DBs and the performance sold us, plus gave us natural language text searches

Phase 2: Final Design

Front End

Request Handler

OODT

Cluster Scheduling System

TAP Service

SOLR

NRAO MD DB & obscore view

ALMA MD DB & obscore view

ALMA NGAS

NRAO NGAS

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

www.nrao.edu • science.nrao.edu