End to End Operations at the National Radio Astronomy Observatory

Nicole M. Radziwill
National Radio Astronomy Observatory, 520 Edgemont Rd., Charlottesville, VA, 22903

ABSTRACT

In 2006 NRAO launched a formal organization, the Office of End to End Operations (OEO), to broaden access to its instruments (VLA/EVLA, VLBA, GBT and ALMA) in the most cost-effective ways possible. The VLA, VLBA and GBT are mature instruments, and the EVLA and ALMA are currently under construction, which presents unique challenges for integrating software across the Observatory. This article 1) provides a survey of the new developments over the past year, and those planned for the next year, 2) describes the business model used to deliver many of these services, and 3) discusses the management models being applied to ensure continuous innovation in operations, while preserving the flexibility and autonomy of telescope software development groups.

Keywords: end to end, operations, archive, proposal, data processing, integrated

1. INTRODUCTION

The Office of End to End Operations (OEO) aims to broaden access to NRAO facilities by simplifying the observing process spanning NRAO telescopes, and promoting the "One Observatory" synergy between NRAO instruments from the perspective of the research scientist. The process addresses proposal preparation and submission, proposal evaluation, observation preparation and monitoring, automated (pipeline) data reduction, permanent data archives, and archive retrieval and analysis facilities. OEO has also been involved in software development for imaging, image display and analysis, although its role in the future may be focused more on deploying these systems for use as archive services instead of developing the core packages.

OEO focuses on successful technology transfer from development to operations, taking an integrated approach that recognizes the dynamic interactions between researchers and the systems they use for scientific investigation. OEO duties include management, maintenance, and incremental development of the following Observatory-Wide science support systems with stakeholders at NRAO telescopes.

- The Data Processing environment, including Common Astronomy Software Applications (CASA), prototypes applying high-performance computing, and involvement in GBTIDL/K-band focal plane array pipeline development with Green Bank staff;
- The Proposal Submission & Management System accessed through Interactive Services, which provides facilities for astronomers to submit proposals to NRAO, and additional functions to process, handle, and referee those proposals;
- The NRAO Data Vault at http://archive.nrao.edu and http://archive.cv.nrao.edu (which provides users with the ability to review and test new services in beta before they are formally released), a web-based platform for accessing historical observations and obtaining proprietary data, coupled with a distributed information infrastructure for long-term data curation,
- Google Sky, to make visual searching and cone-searching an integral part of traditional archive searches on radio data, as well as to incorporate Virtual Observatory services more intuitively, and to make it easier for scientists to promote their research to the general public through Education and Public Outreach (EPO),
- NRAO Interactive Services at http://my.nrao.edu - a dashboard for astronomers using the research facilities to manage their personal profile, data, proposals, and observing schedule; additionally, to gain access to documentation and helpdesk functions

Observatory Operations: Strategies, Processes, and Systems II, edited by Roger J. Brissenden,
David R. Silva, Proc. of SPIE Vol. 7016, 701612, (2008) · 0277-786X/08/$18 · doi: 10.1117/12.788467

With the exception of the data processing environment, this article will present the current status of developments within NRAO OEO, and summarize lessons learned and directions for the future.

1.1 Archive and Pipeline Operations

There are many people around NRAO working on archive operations, and the role of OEO is to monitor progress in all these areas and gradually form a cohesive view of the archive system across NRAO. The European arm of ALMA computing, for the ALMA construction project, is wrapping up development of their archive infrastructure and interfaces based on the NGAS technology developed by ESO over the next few years. Since 2003, NRAO the New Mexico operations group has pioneered the initial prototyping of an online archive for the Observatory, and since 2005 has dedicated effort to integrating the NGAS system into archive operations. Green Bank operates a backup system for its data, but has focused its software development in areas other than archive interfaces.

Because these projects are all operating according to different requirements and on different timescales, the challenge of providing a "One Observatory" archive infrastructure and interface has been substantial. To accommodate these differences, EVLA has adopted the NGAS technology and conducted an initial implementation in operations. Feedback from this experience was provided to the international ALMA archive group, and additional enhancements were made within the project to address these changes, strengthening the archive deliverable for ALMA as well. This information will be used to support development of the NAASC proto-archive as well. The exchange of information has been extremely beneficial, even given that the projects are well underway and the implementations becoming more rigid.

Currently, OEO staff are working on building out the GBT/12m/140ft archive node in Charlottesville, enhancing the VLA/VLBA archive in Socorro so that it is able to support EVLA data archiving, and working to build out the NAASC proto-archive in Charlottesville. A major priority over the next year will be to improve NRAO's ability to manage pulsar archiving, especially for the GBT where the data have not been routinely archived for the long term.

1.2 Proposal Operations

The NRAO Office of End to End Operations also provides software and systems support for three proposal deadlines each year (February 1, June 1 and October 1). The NRAO proposal submission system was originally developed by EVLA computing, and following their lead, the system today (which has been fully transitioned to OEO) is web-based. As of early 2007, NRAO's proposal submission services are fully outsourced to Open Sky Software (http://www.openskysoftware.com) in Austin, Texas. NRAO selected Open Sky not for its astronomically influenced company name, but because several of its owners and staff members are either PhDs or have advanced degrees in physics. As a result, they not only have a keen understanding of the academic culture but also are familiar with the operations of observatories.

Partnering with Open Sky has enabled NRAO to make several significant changes to its proposal submission tool (PST) over the last two years. The PST is now a part of NRAO Interactive Services at http://my.nrao.edu with a new, more intuitive user interface and improved performance. A view of the PST utility within Interactive Services is shown in Figure 1.

1.3 Data Processing Operations

Over the past year, development of the Common Astronomy Software Applications (CASA) interferometry package has been in the domain of OEO. The Common Astronomy Software Applications (CASA) package is a suite of applications for the reduction and analysis of radio-astronomical data. The interferometric computations in CASA are derived from the former AIPS++ software, with the single dish portions derived from the Parkes ASAP package. Algorithms are written in C++. The old interface has been replaced by a powerful, well known, and widely used scripting language (Python). It uses Open Source libraries for MATLAB-style plotting (matplotlib), and other graphical functions such as Qt. The beta version of CASA was released in the fall of 2007, and under interim manager Brian Glendenning, a public version of the beta is being made available in June 2008 for testing by a wider audience. CASA tutorials will be presented at the NRAO 2008 synthesis summer school as well.

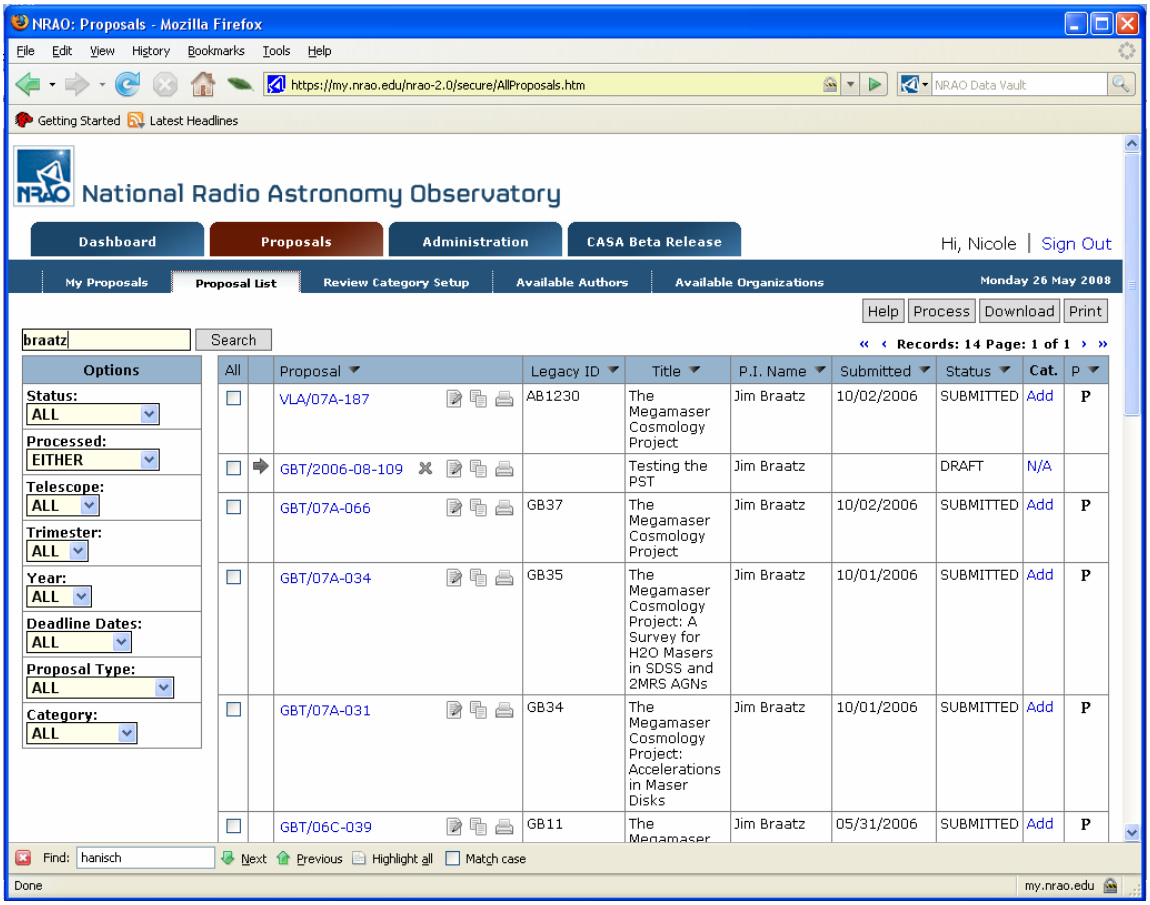

Fig. 1. NRAO Proposal Submission Tool v2008.2 showing all proposals given a specified search term.

1.4 Virtual Astronomical Observatory (VAO) Operations

NRAO is a partner in the transition of the National Virtual Observatory (NVO) to the Virtual Astronomical Observatory (VAO), and participated in developing the proposal for VAO operations that was recently submitted to the NSF and NASA. The NVO project is now completing its final year, and plans for the operational Virtual Astronomical Observatory (VAO) are being made as NRAO participates in proposal development with the project team. NRAO participation in the NVO technical effort, led by D. Tody, has emphasized participation in ongoing IVOA standards development and related implementations, and producing and publishing VO-ready data from NRAO instrumentation. In standards development, the spectral access protocol and Spectrum data model now have multiple implementations. Progress has been made on specification of the new Table Access Protocol (TAP), and on the Simple Image Access (SIA) V2 design, both of which will be a major topic of development for the next year. Work has also addressed a scalable applications framework, which potentially will involve a joint US-EU (OPTICON) implementation effort. Continued effort in this area is dependent upon the outcome from the VAO proposal process in the spring and summer of 2008.

2. RESPONSIBILITIES WITHIN THE OBSERVATORY

In addition to promoting NRAO's "One Observatory" model, OEO is accountable to telescope operations to provide services and deliverables that are either shared across the Observatory, or provide broad benefits that can be extended to other areas. The following sections describe the current and anticipated expectations these divisions have for OEO over the next 18-24 months.

2.1 New Mexico Operations/EVLA Construction

OEO supported the proposal submission software for VLA throughout 2007 and will also begin supporting electronic submission of VLBA proposals in 2008. In addition to supporting regular upgrades and enhancements, OEO will ensure that proposals can be submitted without delay as new EVLA functionality comes online, including various receiver bands and WIDAR functionality for resident and open shared risk observing. OEO will also upgrade the proposal management system for EVLA when the scientific staff identifies the new processes they wish to use. OEO will also ensure that data from the EVLA WIDAR correlator can be ingested into analysis packages during testing and commissioning, and will focus on developing techniques to effectively handle large (500GB-5TB) UV datasets in the data processing packages. EVLA (like NAASC) is also relying on involvement from OEO and CIS to help establish its archive and data handling infrastructure to manage EVLA data. Another critical expectation from EVLA is that OEO will manage the process of algorithm development to ensure that the promise of EVLA can be realized, although at present there are no resources in place which means this cannot be guaranteed.

2.2 Green Bank Operations

In addition to maintaining and upgrade the proposal submission system as new capabilities become available for the GBT, OEO has supported the GBT Dynamic Scheduling System (DSS) and the development of new instrumentation for the GBT by providing high-performance computing solutions over the past year. For the DSS, OEO staff devised and implemented the scheduling algorithms to sort observing sessions ranked by a separate algorithm, and provided leadership implementing a synchronization service to provide updates on portable devices such as the iPod and iPhone. Over the next year, OEO will support the release of the GBT DSS and will also modify the proposal submission process to align with dynamic scheduling. With respect to instrument development, OEO resources have played a key role in software engineering and hardware integration for the new Green Bank Ultimate Pulsar Processing Instrument (GUPPI), which is based on FPGA technology. This involvement will continue as the instrument is commissioned and made available to users. OEO staff will also support pipeline development for the K-band focal plane array as needed.

2.3 North American ALMA Science Center (NAASC)

The key concerns for the NAASC at present are CASA development and ramping up user support for the package, developing a helpdesk which functions well in the international context, and (like EVLA) establishing an archive and data handling infrastructure which is compatible with similar systems across NRAO. OEO and CIS are collaborating with each other and with external facilities like the National Center for Supercomputer Applications (NCSA) to ensure that the structure is as cost-effective as possible.

2.4 Education and Public Outreach (EPO)

Just as it is important for NRAO to clearly communicate the value of its complementary suite of instruments to the general public, we must also communicate the same message to current and potential observers. In addition, NRAO can benefit from using the channels and methods of communication used by EPO to communicate with its audiences (scientists, the media, teachers and students, and the public). To achieve these goals, OEO works with EPO in three areas: presentation of data from NRAO archives to the public through applications like Google Sky, communicating the NRAO message through its web site, and technically supporting new media applications such as social networking for outreach that will include the scientific community. Over the past year OEO has played a key role in the development of the new NRAO website.

3. SURVEY OF DEVELOPMENTS

The following sections provide a summary of some of the work that has been done as part of OEO since 2006.

3.1 NRAO VLA Archive Survey (NVAS) Project

The NRAO VLA Archive Survey (NVAS) project, which is the primary subject of another presentation for SPIE 2008, was initiated by NRAO scientist Lorant Sjouwerman. His goal was to build a pipeline using AIPS that could be used to process historical VLA data and generate continuum images. This effort was very successful, and coupled with the work of Jared Crossley who led the historical processing and verified image quality under the guidance of Sjouwerman, there are now over 72,000 images available. These images have been published as collections in the Virtual Observatory, and are also accessible through the Data Vault at http://archive.nrao.edu.

3.2 Data Vault

The NRAO science data archive has expanded over the past few years to provide convenient access to science data products from the web. Over the past year, OEO has been actively renovating the web interfaces to the archive to provide greater stability, increase scalability of user access, improve ease-of-use, expand the collection of science products, provide web service APIs for external tools such as the VAO, integrate value-added services such as Google Sky, and ultimately provide a fully featured and modern web experience to NRAO archive users. Over the past two years substantial effort has gone into making archive search functionality simpler, easier to use, and much quicker. The new archive interfaces, collectively known as the NRAO Data Vault, currently provide access to VLA, VLBA, and GBT raw data and over 125,000 new calibrated VLA and VLBA images. A view of the main Data Vault page is shown in Figure 2, and the typical view of search results is illustrated in Figure 3.

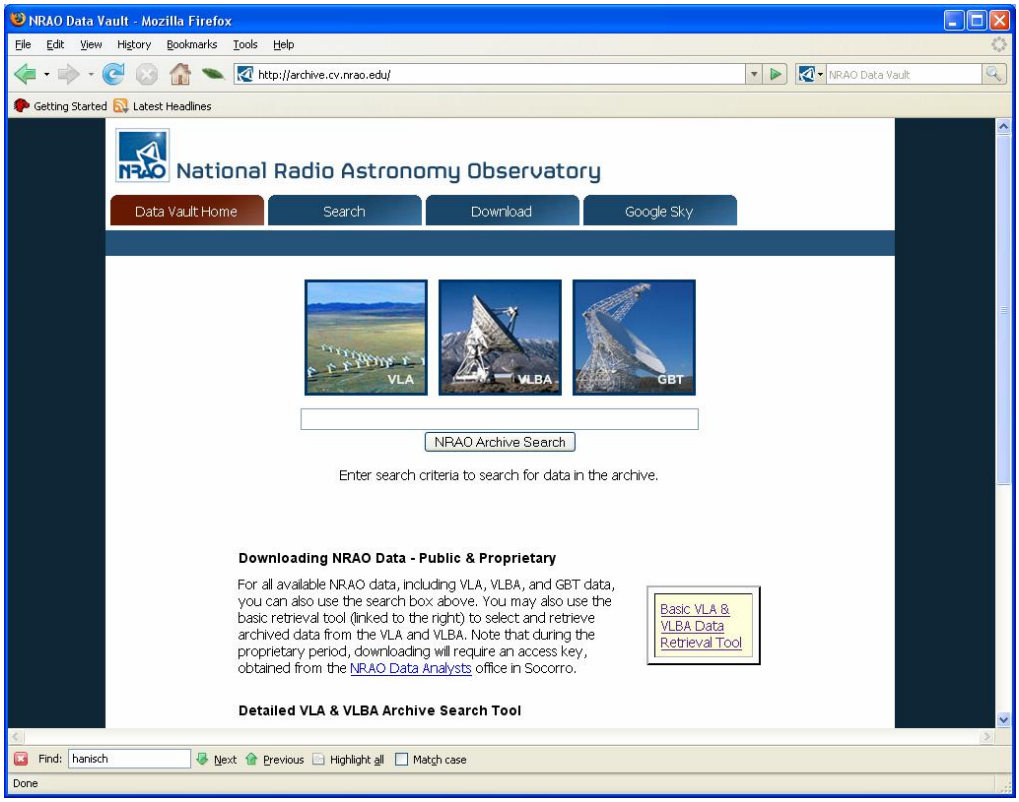

Fig. 2. Main page of the NRAO Data Vault at http://archive.nrao.edu.

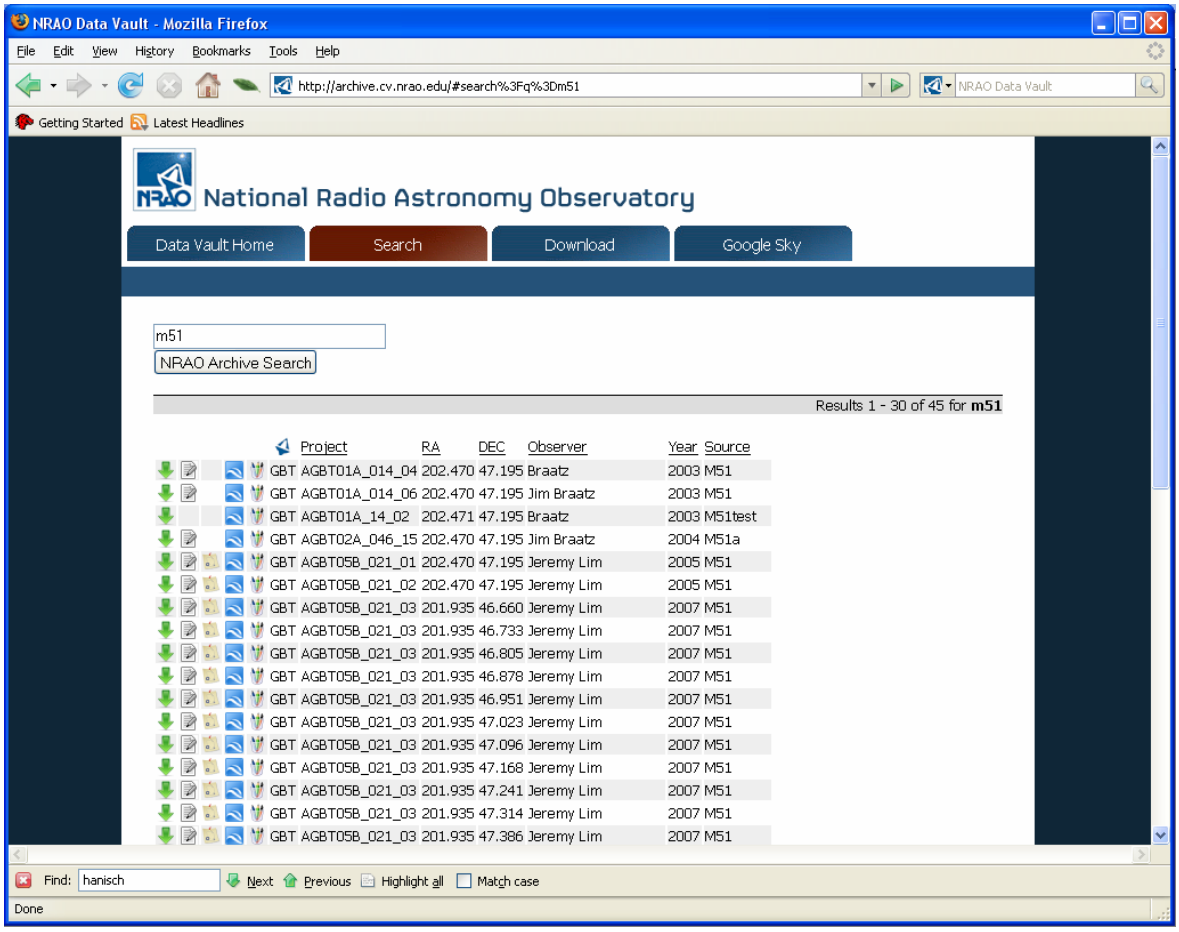

Fig. 3. Typical search results display.

A major part of building out the Data Vault is expanding the collection of pipeline-processed images and maps that are available. Since late 2006, the VLA pipeline developed in AIPS by L. Sjouwerman has been heavily applied to archived VLA data, resulting in over 72,000 continuum images that are now accessible from the archive. With continued effort, the entire VLA archive in all array configurations will be processed by the end of 2010. We plan to do the same re-processing of archive data to general spectral line maps using pipelines under development by J. Uson, B. Cotton and L. Sjouwerman starting at the end of 2008, pending continued availability of their effort.

The development of an archive infrastructure that addresses the operations needs for all NRAO telescopes has also been an area of activity since 2006. At present, we have a draft architecture complete for the EVLA implementation, and we are now identifying how that needs to adapt to support ALMA and GBT needs and the requirement to easily locate data in one search regardless of what NRAO instrument produced that data.

OEO also plans to work on developing services for surveys and large projects observed on NRAO telescopes, including web/archive hosting and facilities for storing post-processed data products. Addition and integration of data collections from the 12m and 140ft telescopes will also begin in the summer of 2008, at the same time that we begin developing "semantic searches" where search strings such as "molecular clouds" will yield useful results.

3.3 NRAO and Google Sky

Google Sky (http://earth.google.com/sky), a networked application based on Google Earth (http://earth.google.com), enables visual browsing and searching of the astronomical sky. Since its debut in August 2007, Google estimates that over 4 million users worldwide have downloaded and now use the application. In May 2008 the web-based version of Google Sky was integrated into the NRAO Data Vault to provide a visual complement to the text-based search. This is shown in Figure 4.

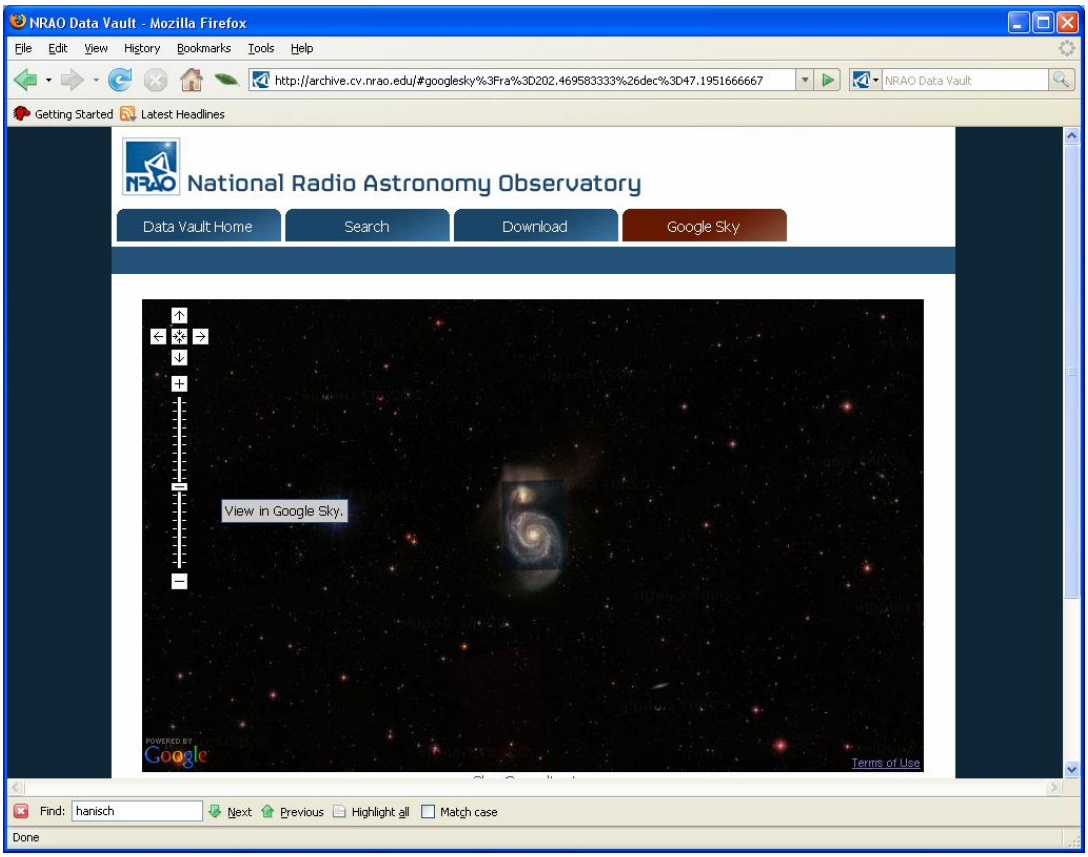

Fig. 4. Clicking on a search result from the Data Vault will show the optical view of the sky in that region.

To promote awareness of radio data among the general population, OEO constructed a set of 6 "radio skymarks" for Google Sky which were demonstrated at the AAS meeting in January 2008. These are available at http://archive.cv.nrao.edu/skymark/. NRAO also submitted its work to the Sky Gallery which provides broader access to the NRAO collection of radio skymarks. (http://earth.google.com/gallery/kml_listing.html#csky#s1#e10 or click on "KML Gallery" from http://earth.google.com/sky/skyedu.html).

NRAO continues to work with the Google Sky development team in Pittsburgh to help make the application more useful to the astronomical community. In the summer of 2008, the "NRAO Showcase" will be developed which will be similar to showcases in place for Chandra, Hubble and Galex, which will highlight approximately 30 phenomena and will be translated into approximately 20 languages. Later in the year, we will construct a pipeline to translate FITS images to the KML required by Google Sky, with the goal of developing a browsable radio background using NVSS and VLSS survey data. By the beginning of 2009, we also plan to have prototypes of the 3D structures of NRAO telescopes available in Google Earth linked to real-time views of what's being observed.

3.4 NRAO Interactive Services

NRAO Interactive Services launched in mid-2007 at http://my.nrao.edu. The goal of NRAO Interactive Services is to provide one online location where users of NRAO telescopes can log in to find their proprietary data, communicate with their collaborators, build proposals, and otherwise find resources that they need to obtain NRAO data and do science with it. At present, this only provides a portal for researchers to find resources, but the intent is to create a platform for each of the NRAO sites and telescopes to be able to easily plug services into the framework.

3.5 Proposal Submission & Management Utilities

The PST database has been re-factored with additional validation rules, improving the overall integrity of the submission process. The technology is fully outsourced and costs less to support on an annual basis than the previous in-house arrangement. The external developers have become a part of the core OEO technical development team and work effectively with NRAO staff.

For the June 2008 proposal deadline, the VLBA (including proposals for HSA and global mm VLBI) has been added to the PST. Later this year, development plans include a) allowing sites to configure telescope-specific resource data, eliminating the need for software engineering support every time an instrument is updated, b) adding a new class of users, referee users, to improve the efficiency of the evaluation of proposals, and c) providing data to users that can be utilized during the observation stage (e.g., providing users with source catalogs from the PST to import directly into observing programs). The PST will evolve as the dynamic scheduling systems at the EVLA and GBT are developed, and will continue to respond to new instrumentation for all NRAO telescopes. Another important addition which will be addressed in 2009 is updating the EVLA proposal process so that proposals are defined in terms of science objectives rather than instrumental specifications, which is the mode being developed for ALMA.

3.6 Plans for 2008 and 2009

The priorities for development are derived from online surveys, linked to NRAO Interactive Services at http://my.nrao.edu and the Data Vault archive interfaces at http://archive.nrao.edu. Based on this input from the user community, supplemented by input from the strategic planning process at NRAO, OEO plans to a) host web and storage solutions for the processed data from its "large" projects with observing time that exceeds 200 hours, b) find a solution for pulsar archiving to preserve this data more effectively long-term, c) integrate additional VO services into the Data Vault, d) provide Google Sky with a radio background using the NRAO VLA Sky Survey (NVSS) and the VLA Low-Frequency Sky Survey (VLSS), and e) implement a visual cone search using Google Sky.

4. INNOVATION NETWORKS AS A MANAGEMENT MODEL

According to Cross & Parker [1], work in an organization is more likely to get done as a result of the informal connections between people who work together well and trust one another (the "informal networks") than through the formal, hierarchical chain of command. This revelation becomes particularly apparent when the people within an organization who *must* work together on similar or connected software systems are managed by different bosses, and must accommodate different sets of requirements and different timescales for implementation. The solution to this problem is to recognize that people will take advantage of cross-organization connections *if those connections help them achieve their own goals*, and if those connections do not create a conflict of interest with their other perceived obligations. This is a very simple conclusion, but one that can often take months or years to recognize in a situation where disparate workgroups must be brought together to achieve a common goal.

Two examples of how networks are being cultivated at NRAO are for archive development and to catalyze the implementation of high-performance computing techniques:

- **Archive Development Network**: The NRAO archive is developed by several people across the Observatory who are each responsible for meeting the requirements of one telescope. As a result, attempting to manage these efforts deterministically, as a project, was not fruitful. However, once a network mindset was adopted, development flowed much more smoothly according to a service-oriented software model where anyone could build a service to attach to the archive.
- **High Performance Computing Network:** In 2007, OEO sent a group of observatory software engineers to the Supercomputing '07 conference which is the premier venue for experts in emerging solutions to the most challenging data management problems. This helped form discussion groups within NRAO on these topics which are now very active. OEO will continue to sponsor attendance at the Supercomputing conferences held each fall, and plans to launch a regular internal newsletter to help the scientists, software engineers and computer engineers working on these problems to communicate their results more effectively.

This network-style management model was further illuminated in early 2008, when NRAO more aggressively started to explore options for an "Integrated Science Center". As of this time, operations for the VLA, VLBA and EVLA are centered in Socorro, New Mexico, and operations for the GBT are based out of Green Bank, West Virginia. The North American ALMA Science Center (NAASC) is headquartered at NRAO in Charlottesville, Virginia. Various review committees had suggested that NRAO explore different operational models, and so several options (including geographically co-locating all science support) were considered. The outcome of this exercise was recognizing that as a telescope is built and commissioned, expert astronomers and engineers must be involved in all aspects of systems development but as knowledge about those systems increases, maintenance and continued enhancement can be done increasingly by generalists whose knowledge is more topical but spans more telescopes. Following the diagram in Figure 5, new software and systems should be developed by the generalists in the leftmost telescope-specific organizational units. As more and more is learned about those systems, organizational units that involve more generalists (to the right) can be leveraged. The implication is that it is more effective to develop specific systems that have the potential to be generalized, rather than attempting to initially develop a general system that must be fit to the needs of a telescope.

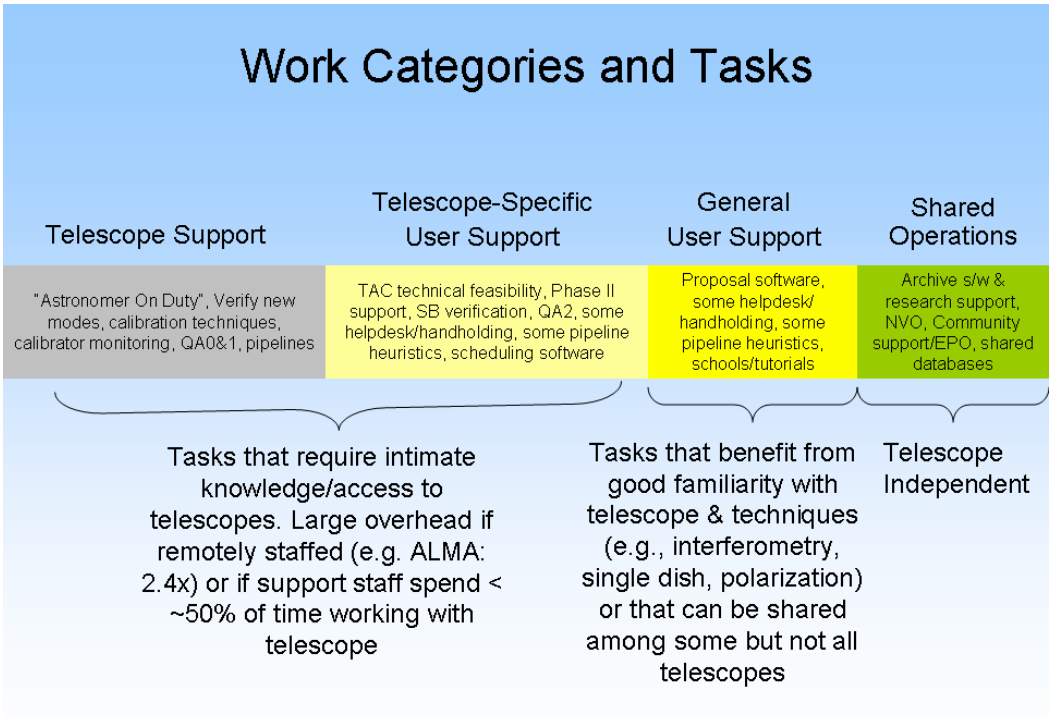

Fig. 5. A process-based model for integrating science operations across NRAO. (Diagram by J. Hibbard)

5. CONCLUSIONS AND LESSONS LEARNED

Several lessons have been learned over the past two years that NRAO has treated "end to end" systems development in the operational context instead of as a unique development project. These conclusions are not intended to be exhaustive or prescriptive, but instead to reflect the lessons learned throughout one observatory's journey of trying to steer its software development organizations toward a sensible level of cohesion at the observatory level, while preserving autonomy at the telescope level.

For example:

- An observatory-wide software development or "end to end" effort should provide a means for scientists pursuing their individual interests to contribute their independently derived tools, services and artifacts for the greater good of the user community. The NVAS project described earlier was an example of a promising project for which a small amount of additional effort and resources were able to yield a huge long-term benefit.
- The outcome of commonality in software, systems and data models across telescopes is due to the individual efforts of people in key roles who believe that such commonality is beneficial. No amount of management or decree will motivate groups to develop common systems unless there are clear benefits for all involved, and these are understood by most members of a workgroup.
- Outsourcing is a promising arrangement for software systems that are telescope-independent. Use of contractors enforced some discipline in our development process (e.g. if a scientific sponsor does not get requirements or reviews in by a required deadline, the software development to realize those requirements simply will not happen). Additionally, contracting development on a cycle-by-cycle basis enables management to avoid cost overruns (it is always possible to skip new development for one cycle to balance the budget).
- There is a continuum between "end to end" systems and the education and public outreach efforts of observatories that may not be emphasized in many institutions. NRAO has chosen to embrace this through its efforts to support Google Sky development for the radio community, and to make it very easy for science done with NRAO telescopes to gain an audience through Google Sky.
- Requirements and expressed needs are very different when they are solicited from outside the observatory, than when requirements are collected within the observatory. A balanced approach to requirements gathering could be helpful to capture the needs of those less familiar with a telescope, while integrating the deep knowledge held by the scientists and engineers close to a telescope.

These conclusions are, undoubtedly, a function of the NRAO's financial context and organizational culture. However, the notion of cultivating a network for the purpose of speeding development and innovation, particularly in the software area, is promising. The network model should be investigated by any manager whose goal is to bring together, in a mutually beneficial and productive way, workgroups that have historically operated independently. The opportunities are plentiful for new organizational models in astronomy that support order and individual innovation simultaneously.

ACKNOWLEDGEMENTS

The development of the emerging end to end subsystems (archive system, proposal system, and pipeline installations) at NRAO would not be possible without the contributions from many people, inside and outside the Observatory. The author would like to acknowledge these contributors, including John Benson, Lorant Sjouwerman, Jared Crossley, Ron DuPlain, Stephan Witz, Susan Loveland, Bryan Butler, Doug Tody, Dana Balser, Joan Wrobel, Jean George and Ashish Arte at Open Sky, and Eric Sessoms. Additionally, there are many scientists outside the Observatory whose interactions with OEO through surveys and emails have illuminated the way forward. There are undoubtedly many more people behind the scenes whose efforts have contributed, directly and indirectly, to the state of the system as it exists today and for this the development teams are also grateful.

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

REFERENCES

[1] Cross, R. & Parker, A., [The Hidden Power of Social Networks], Harvard Business School Press, Cambridge, MA (2004).