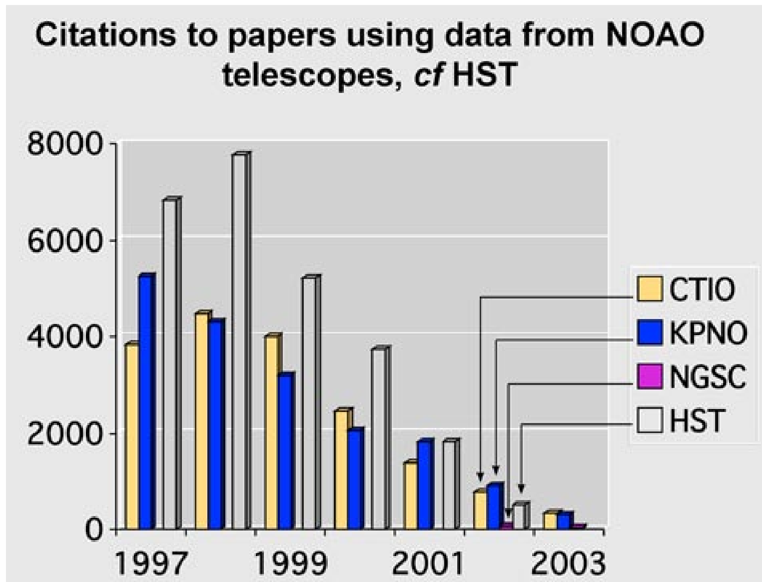
The NSF Senior Review

Jeremy Mould

When I was a postdoc, the solar astronomer Ron Giovanelli asked me why I worked in Tucson. As a frequent visitor to the Solar Division, the forerunner of the National Solar Observatory, he had his own reasons.

"If it can be observed," I replied, "it can be observed at Kitt Peak." He laughed at my youthful bravado, and remarked only that it was a strong statement. In those days, the Uhuru X-ray satellite was the only observatory in space. Nowadays, NOAO would construct a more carefully worded statement, but the national observatory has similar aspirations, to provide peer-reviewed access to the finest optical/infrared (O/IR) facilities in both hemispheres for observations best done from the ground.

Reviewing our array of instruments on pages 24–27 of this newsletter, I feel that this is a legitimate claim. It will be defended vigorously at the pending senior review of the NSF's astronomy facilities for all wavelengths, as announced in January at the American Astronomical Society meeting in San Diego.



ADS citations evaluated in 2004 of papers published in the year denoted on the time axis. Naturally, papers take time to attract citations, but flatten out after a few years. HST data are from the STScI Newsletter *volume 20, issue 2.*

NOAO's 4-meter telescopes provide the best deep survey capability around, with the possible exception of the CFHT in the north. Our Gemini spectrographs, GMOS and GNIRS, provide the best low-resolution capability at the low multi-object level. We are more patchy in high-resolution spectroscopy, with only the classical Echelle and access to HIRES in the north, plus the resurgent Phoenix, the soon-to-be-commissioned bHROS on Gemini South, and future SIFS on SOAR. TSIP provides limited access to high multi-object spectroscopy at the MMT, Keck, and Magellan. T-ReCS and Michelle complement Spitzer with an order of magnitude higher angular resolution on bright thermal IR sources. Coming down the pike, we have NEWFIRM and (thanks to Fermilab) the Dark Energy Camera retaining the lead in survey astronomy. Of course, the NSF astronomy division senior review is intended to find NSF resources for design and development of three vital facilities for the future, the Large Synoptic Survey Telescope, the Giant Segmented Mirror Telescope, and the National Virtual Observatory. And so we support the senior review.

But NOAO's case for O/IR growth from the senior review has three sides to it. The first is the quality of the facilities, just described. The second is demand, which you can gauge from page 23 of this newsletter. And the third is outcomes. In Astrophysics Data System (ADS) citations, if one accepts the Hubble Space Telescope (HST) as the gold standard for observatories, NOAO performs very creditably (see figure).

My point here is not to argue about how many ground-based telescopes it takes to equal a NASA Great Observatory, but rather to assert that retaining and developing NOAO's current facilities is probably as scientifically important as retaining and developing HST, for the time being. Our other key metric is the number of thesis students NOAO supports annually, and this too is a significant portion of the national total.

A further community-based input to the senior review that NOAO is supporting is the roadmap, or Long Range Plan, for national O/IR facilities, which is currently under development by a committee chaired by Caty Pilachowski of Indiana University. Details about this roadmap should be available in the next NOAO-NSO *Newsletter*. Progress to date by the committee is charted at *www.noao.edu/dir/lrplan/lrp-committee.html*, and your input is requested.

Jeffrey Kantor

LSST Data Management Project Manager

Jeffrey Kantor is the project manager for data management for the Large Synoptic Survey Telescope (LSST). Kantor joined the LSST effort in June 2004. He brings a diverse background in information technology to the project, having worked for a variety of organizations in industry, NASA, and the US Department of Defense, in areas ranging from aerospace to semiconductor manufacturing to consumer goods. Away from work, he enjoys spending time with his family, and he hopes to find time in the future to return to soccer coaching and refereeing.

Q. What are your major areas of responsibility for the LSST?

My team is responsible for everything that happens to LSST data once it leaves the detector, from acquisition to processing, pipelines, archives, and making the data available to the community.

The LSST is somewhat unusual in that a lot of the project's data management development is being planned up front; it tends to be an add-on in other projects that I have worked on. The LSST project directors recognize that, given the huge data volume and the related processing requirements, the data management job would not get done if we waited. It is truly one of the more challenging parts of the project.

Q. What are some useful analogies for the LSST data management challenge?

The LSST is really several orders of magnitude beyond previous sky surveys; we will be producing a volume equivalent to the entire Sloan Digital Sky Survey every night. The LSST will generate 15 terabytes of raw images per night, which translates into approximately 30 terabytes of processed data. You can think of it as stamping out a new DVD of data every 10 seconds! That's what happens when you operate a three gigapixel camera all night, every clear night.

Q. What are your priorities at this point?

Our number one priority is obtaining sufficient funding, starting with an effective design and development (D&D) phase proposal, which has been submitted to the NSF. Our second priority is to build a strong data management team. We have now had two team meetings, and a science meeting, which helped produce three very capable research teams. We identified the critical issues and solicited participants, then put plans together for the next two to three years. At the end of that, we will produce a construction phase proposal to cover the next five years.

There are three vertical layers in our project: applications (which covers archives, pipelines, and tools), middleware (which includes databases, GRID technology, and software engineering tools), and infrastructure (such as computers, communications, and storage.) Each of the three teams has a specialty in one of those three areas.

Horizontally across the teams, we are looking at what we need to do to get through the basic data processing with sufficient speed and accuracy to issue real-time alerts on transient objects, which needs to happen fast. Things like archiving and the weak lensing data pipeline can happen at a more conventional pace. Finally, we are looking at public access tools and protocols.

Q. How have you broken down these complex tasks into manageable pieces?

We're taking multiple approaches to some of these challenges; that is what the D&D phase is all about. Some of this involves trade studies, while in other areas we are creating multiple prototypes using precursor data from a half-dozen other surveys.

One true research area is estimation of the point spread function (PSF). We are trying to achieve pretty high astrometric and photometric accuracies with LSST observations, and there are different ways to calculate the PSF, especially from stacked images.

We will need 50 to 100 teraflops of computing power. There is a real issue regarding how much processing you should do close to the observatory versus the need to ship the data somewhere else first. The question is: do we move the data to the processing or move the processing to the data? That is where GRID technology helps.

continued

Q&A continued

We're also doing a large number of technological trades, such as the strengths of a relational database versus an object-oriented one compared to distributed file systems. Another question is archiving on disks versus tape.

We want to make our technological selections at the end of the D&D phase to take maximum advantage of the positive technological trends in CPU power, disk storage, and networking. These trends suggest that hardware capability and cost should not be a limiting factor by the era of LSST operations in 2012 and beyond.

Q. Who are your current partners?
We have a number of institutional partners across the three teams. The "Eastern" team is Harvard, Brookhaven National Laboratory, Princeton, and the University of Washington (we sort of tagged them geographically); Johns Hopkins University just joined LSST and will probably join this team. The "Central" team consists of NOAO, the National Supercomputing Center at the University of Illinois, the University of Arizona, and the University of Pittsburgh. The "Western" team is the University of California-Davis, Livermore National Laboratory, the Stanford Linear Accelerator Center, and, very shortly, the San Diego Supercomputer Center.

The NCSA and possibly the San Diego Supercomputer Center will be the hosts of the data, which makes sense since they are NSF-funded facilities and we will be significantly funded by the NSF, along with the DOE and private institutions.

Microsoft and Google are acting as independent advisors, at a consulting level of reviewing our plans for certain areas and providing comments. They have not gone beyond that yet, though there's a possibility that their roles could grow. They recognize the size of the data management challenge that we have, and it relates very closely to problems that they have attacked before. We are also in discussions of a similar nature with IBM.

Q. What attracted you to the LSST project?
I've always been fascinated by space and learning more about the cosmos. I did work before with NASA and the International Space Station, and the majority of my career has been in space and science. I tried the business world, and found that I like this more. I'm very pleased with life in Tucson, it's a great town, a good size, and it has a tremendous amount to offer. I certainly enjoy the warm weather.

Q. What sorts of secondary spin-offs might emerge from the LSST data management challenge?
We anticipate that we'll come up with some fairly innovative ways to manage large databases. Some of this will be "provenance" — all those things that allow you to assess the quality of the data after it is taken, I expect breakthroughs there. We don't want to double or triple the size of the data set in the process of doing that, and this is very applicable to other problems.

The fact that the data will be publicly available all the way down to the K-12 education level is fantastic. There will be lots of things that can be done with the data from a public outreach standpoint. Having the whole Universe accessible in one database will enable teachers, students, and amateur astronomers to ask and answer lots of new questions. For example, I can envision an "LSST@Home" type of application for amateur follow-up of transient object alerts, helping us decide whether they are "real" or not.

Q. How can the astronomical community get more involved in helping guide the development of the LSST data management system?
One way is through the "Science" page of the project Web site at *www.lsst.org*. It now has a discussion forum for community science participation. We hope that this will enable the community to suggest ways to exploit the data, and to help us feed those desires back into system requirements. Now is the time to get involved.

Notable Quotes

"Your Nebraskan pragmatism—and your knowledge of the magician's tricks—tilted you toward the sciences, especially astronomy. (Maybe this is why the occultists, future predictors, spoon-benders or mind readers on your show never left without having been challenged.)

You were host to writers, children, intellectuals, and nitwits and served them all well, and served the audience by your curiosity and tolerance."

—Excerpt from a posthumous tribute to television talk show host Johnny Carson by comedian and author Steve Martin in the New York Times*, 25 January 2005. One of Carson's most frequent guests during his 30-year run hosting "The Tonight Show" was space scientist Carl Sagan, who appeared more than two dozen times.*