

Discussion Summary: Astronomy's Changing Scientific Culture

The first discussion session of day 2 was entitled “Astronomy's Changing Scientific Culture” and was devoted to four rather broad topics - these include the concept of public-private partnership observatories, astronomical surveys, data products and pipelines, and the concept of a national virtual observatory. The session moderator, Ata Sarajedini from the University of Florida, provided a brief introduction after which Alan Dressler of the Carnegie Observatories presented a historical overview of the ways in which public and private funds have partnered to expedite the astronomical research endeavor.

Historical Overview

Dressler began by noting that in the United States, there is a long and rich history of public and private institutions funding the establishment and operation of astronomical observatories. These efforts go back to the end of the 19th century when many of the elite private colleges in the US had built observatories, most notably Harvard and Yale, but public institutions, for example, the University of Chicago and the University of California, had also invested in major research facilities. In 1904 George Ellery Hale — fresh from his accomplishment of constructing the world's largest telescope at the U. of Chicago's Yerkes Observatory — convinced trustees of the newly founded Carnegie Institution to put a small part of Andrew Carnegie's fortune to work in the interests of astronomy, which quickly led to two “world's largest telescopes” at Mt. Wilson. Such efforts grew to dominate the landscape of US astronomy in the first half of the 20th century.

At first, access to each optical observatory was for the most part exclusive to the institution that built and operated it, but in the 1930's collaborations began to appear, like the Carnegie-Caltech building of the Palomar Observatory (Hale Telescope) and the U. of Chicago-Texas collaboration at McDonald Observatory. In 1940, Otto Struve, director of both Yerkes and McDonald, proposed that such collaborations be expanded to include more universities, private and public, so that access to telescopes would become more widespread among US university researchers.

A significant change came in US optical astronomy in the 1950's with the founding of AURA. A large consortia of universities collaborated to build, with Federal funding, optical telescopes that would be available to all. The Kitt Peak and later Cerro Tololo Observatories were built as a parallel set of facilities that, by the 1970's, provided all the scientific capabilities available at the “private/university” observatories. According to Dressler, the founding of AURA was the crucial turning point for US astronomy because the long established optical observatories Lick, Harvard, Mount Palomar, Mount Wilson, Yerkes, McDonald, Steward, etc., were not the agents of expanding the system to provide universal access, the approach Struve seemed to be promoting before World War II. The founding of AURA and the building of what we now call NOAO led to two decades of competition between the private/university observatories and the “national facilities.”

Although very productive from a scientific perspective, Dressler thought that

in some ways the competition between the private/university and national observatories had been quite unproductive. Particularly detrimental for the development of US optical astronomy, he believed, was the fact that the NSF was never able to claim substantial credit for the most powerful astronomy facilities in the world, since most of the “glass” was in the hands of the “privates.” Dressler speculated that this has kept NSF funding for astronomy below what it might have been if the private facilities had offered at least limited access to the whole community over the last several decades.

A series of recommendations starting with the Ground Panel of the 1990 Decadal (Bahcall) Survey, chaired by Steve Strom, continuing with the McCray Committee of the 1990’s and culminating with the OIR from the Ground Panel of the 2000 Decadal Survey chaired by Alan Dressler and Todd Boroson, all focused on encouraging collaboration between the private/university and public observatories. Throughout this time, NSF had made it clear that major investments in the private/university facilities would require broad access to these facilities. The OIR Panel of the 2000 Decadal Survey proposed a modified instrumentation funding in-exchange-for-observing time program, TSIP, which has resulted in the construction of several major instruments and the addition of 75+ nights per year on Keck, Magellan, etc.

The OIR Panel also recommended LSST as the first dedicated time-sampling telescope. The Panel imagined LSST to be a NOAO project, which would provide unique survey data for the entire US astronomical community. LSST has itself become, in fact, a public/private partnership. Finally, the construction of a 30-m class telescope was proposed to be a private/public partnership, the so-called ‘GSMT’ project. The Panel’s intent in making all three recommendations — a package — was to promote a cooperative, non-coercive confederation of all US facilities that would leverage the huge investment of private and university funds with Federal dollars, and capitalize on the strengths of the public (NOAO and Gemini) and private/university observatories. The idea was to build an OIR system that would provide some access to all users to all facilities. Just as important, this would make the NSF a stakeholder in all facilities, not just NOAO or Gemini, so that a case could be made for increased support for ground-based astronomy based on universal access — this was thought to be crucial if the US was to remain competitive with ESO.

Dressler finished up his presentation by presenting some statistics on publications based on observations from the Magellan telescopes. These have been compiled by John Grula of the Carnegie Observatories and are current up to December 31, 2007 (compiled as of February 28, 2008). These statistics strongly suggest that the sharp distinction between public and private access that existed 25 years ago has greatly softened, mainly due to the importance of collaboration in doing major research projects, and the fact the unique facilities like HST drive collaborations that are to a large extent independent of “observatory” boundaries.

Dressler noted that this is a positive trend, because it means that when we add “universal access” to private/university facilities, it is easier to come up with the nights that provide access without collaboration with Consortium members — that is, not as many nights are needed as in the “old days.” It also means that, if we can indeed add that fractional component, say 25% of TMT or GMT or both, and we take into account the collaborations that cross all these institutional boundaries, we really will have achieved,

to all intents and purposes, a system of universal access by US astronomers to all ground-based optical/IR facilities.

The discussion of Dressler's presentation among workshop participants was relatively limited. One member of the audience did note that for a long time, the Palomar Observatory has had a visiting observer program, which is something Dressler did not mention in his presentation. This is clearly an example of the private/university observatories providing observing time to the general astronomical community.

Surveys

The next presenter in this session was Mark Dickinson from NOAO speaking on multi-wavelength multi-collaborator surveys in the era of the GMT/TMT. A survey is defined as a set of astronomical observations that address problems needing large, statistical data sets, often many different coordinated data sets, and uniform data quality. A survey generally produces calibrated, documented data sets for public distribution, which can be used for many purposes, and which are likely to be used by many investigators, beyond the original team that carried out the project. Surveys often, but not always, involve large teams of astronomers. And they typically require substantial resources in the way of telescope time, human effort, and funds.

One way in which the next generation of ELTs such as TMT and GMT will fit into our "survey culture" is that they will frequently be used to follow up objects from surveys that were carried out with other facilities such as LSST, JWST, SPICA, ALMA, to name a few. The TMT and GMT will also be used to carry out their own surveys. The science case documents for the two projects describe large-scale survey projects that could require tens to hundreds of nights of observations.

It is also important to note that TMT/GMT will exist in an era when many other, new, powerful multi-wavelength observatories will be scanning the skies. Astronomers will likely be increasingly comfortable with multi-wavelength science and less likely to be "compartmentalized" into wavelength-specific sub-communities. In this context, the astronomical community will want to review options for multi-facility time allocation in the TMT/GMT era. Multi-observatory proposals are a normal part of the current NOAO time allocation process, and represent a modest but steady fraction of all NOAO proposals. However, there have been very few long-term or multi-cycle time allocations on the 8-m Gemini telescopes. To date, the large (8-m class) telescopes have been excluded from the NOAO Survey Programs, and are only open for "normal" proposals. In contrast, large, multi-season observing programs and surveys are much more common with the VLT, where Large Programs use a significant fraction of available observing time, and even at the US private observatories. If the US astronomical community outside the TMT and GMT partners wishes to be able to carry out survey-type science, a change in this culture of time allocation would be necessary. In turn, surveys should give something back to the broad community in order to make the trade-off between fewer large programs vs. many small programs attractive. Options might include widespread opportunities for community participation in survey teams, and ideally, funding for community data analysis and archival research, much as is done for space astronomy.

Finally, Dickinson pointed out that surveys require archives to distribute their data products to the community. Archives can enable "virtual surveys" of their holdings via

data-mining. Data processing pipelines can facilitate surveys, and since they are so intimately linked, they are the subject of the next two presentations.

The discussion following Dickinson's presentation focused on two particular questions. First, it was pointed out that the follow-up observations that extend and broaden the findings of a survey may require more telescope time than the survey itself. How does one ensure access to both classes of facilities – the survey and follow-up instruments? Second, does public participation in a survey mean access to the observing facility or access to the science enabled by the facility? This latter question led to a discussion of the merits and relative costs of setting up and administering a data archive for a telescope. One member of the audience pointed out that since archives do not contribute to the base operations of a facility, they should be funded by the broader community who will be using them. These topics were delved into more fully in the next presentations.

Data Products

David Silva, currently the Director of NOAO, was the next speaker and he addressed the question of data products and how best to optimize their use. Silva pointed out that the primary data products of an astronomical facility are refereed publications. The secondary data products include raw and/or calibrated observations maintained in an archive with complete supporting information such as appropriate calibration data, software, and quality assessment, to name just a few. Of course, these secondary data products can maximize the number of refereed papers per unit observation.

Producing a useful data archive requires consideration of a number of factors such as expert knowledge of the telescope/instrument system, sufficient data calibration, data quality control and assessment, and timely processing and release of the data in question. On top of this, automatic processing and pipelining requires additional factors to be considered. To this, we undoubtedly need to add the complexity of handling a variety of different types of observations from imaging to long-slit, multi-object, and integral-field spectroscopy. Furthermore, uncertainties due to weather and observing conditions must be considered.

The question of how much priority the TMT/GMT projects should place on archiving and pipelining is one that Silva hoped the audience would discuss. First, Silva suggested that one route we may want to follow to limit the initial costs of operating the archive would be to largely ignore small self-contained datasets, which may have limited appeal to the broader astronomical community. This may 'throw away' the majority of TMT/GMT observations, but it does archive and pipeline-process the observations that hold the greatest potential for data-mining by the community.

The discussion following Silva's presentation focused mainly on the relative merits of archiving and pipelining as compared with the cost and effort involved. The point was made that it seemed a shame not to allocate the required resources to archiving and pipelining after spending the significant funds required to build and operate ELTs such as the TMT and GMT. Audience members also echoed Silva's point that all of the data may not be worth archiving and that it would be more straightforward to concentrate on the fraction of the data that would be most beneficial to the community as a data-mining resource. With regard to funding research using ground-based data, a number of times audience members expressed their frustration that unlike NASA, NSF does not provide

grant money along with observing time in order to reduce and analyze the data. This applies to both observational programs and archival ones.

Virtual Observatory

Alongside the issue of archiving the data from the ELTs is the question of how best to disseminate the data once they have been archived. This is at the heart of the Virtual Observatory (VO) concept, which was the topic of George Djorgovski's presentation. VO is intended to be a complete, distributed, web-based research environment for astronomy with massive and complex data sets. In the U.S., the concept is realized through the Virtual Astrophysical Observatory (VAO; né National Virtual Observatory, or NVO); see <http://us-vo.org>. It is a broad, international initiative, unified through the International Virtual Observatory Alliance (IVOA); see <http://ivoa.net>. VO is a new type of a scientific organization, and an example of a more general trend in all sciences, often described as Cyber-Infrastructure or e-Science.

A key point about the VO is its *systemic nature*: it provides a standardized, open informational and computational environment for both data providers and data consumers, regardless of the scale of the facility, data volume, etc. There are two principal ways in which VO can benefit large telescopes:

1. To facilitate data preservation, by providing a standardized environment for data access and archiving.
2. To optimize their scientific utility: (a) by providing an effective way to select interesting targets from large sky surveys, their fusion, and data mining; and (b) by providing a rich information environment (encompassing both archival data and theoretical simulations) within which one can analyze and interpret the new observations.

There are also some profound sociological and managerial changes brought by the era of exponential data growth. A relatively simple one is that, as the data volumes increase, the value of data ownership decreases, but the value of the expertise needed to interpret the data increases. That should have some tangible implications for various forms of data access rights, beyond the simple proprietary period mechanism, which was really established for the data poverty regime.

The more important challenges manifest themselves in the growing importance and cost of software. As the focus of knowledge discovery shifts from obtaining the data to interpreting the data, the share of the cost and the creativity shifts from hardware to software. Whereas the price/performance of the computing hardware improves exponentially, the relative cost of software development at best remains constant, and probably increases. In data-intensive projects today, software can account for ~30%, and perhaps up to ~80% of the total cost (including everything, from observatory and telescope control and operations, to data processing, archiving, and scientific analysis). This should be recognized and planned for; but our professional culture is still very much hardware-biased. VO can help in some aspects, by providing standardized tools and environments; software development is too expensive and important to afford an unnecessary duplication of efforts.

Related to this is the problem of professional rewards and recognition. Building “software instruments” should be recognized as being at least as hard and valuable as building of telescopes and other hardware. Failing to do so will inevitably result in the brain drain of essential professionals and interested students into other fields, where these issues are recognized and dealt with in a more appropriate manner.

The discussion following Djorgovski’s talk focused on a number of issues having to do with the resources (especially money) needed to set up an archive and then implement the virtual observatory concept. It is anticipated that high-quality fail-safe software will be more of a challenge than hardware, yet it is generally more difficult to obtain private funding for software than hardware. As such, this may be an area where corporations such as Microsoft and Google could be encouraged to contribute to our scientific venture. This would have public relations value for these companies, and it would be a way for them to develop software tools with commercial value at some later time. At the same time, many audience members felt that the best software is usually written by those closest to the telescope/instrument project so it is important to capture what they do and know about the observations to be archived and disseminated.

Summary

In summary, it is the view of the session moderator that the following important points should be gleaned from the presentations and the subsequent discussion. First, we must heed the lessons of history and use them to our advantage. As such, the involvement of the National Science Foundation in the GMT/TMT endeavor is of paramount importance. It is clear that such an involvement promises to be mutually beneficial for the NSF and the US astronomical community. Second, we must continue to expand our efforts in the use of multiwavelength data in order to broaden the discovery potential of the observational material at our disposal. This then leads to the third point, which is the increasing importance of an established and efficiently functioning archive and pipeline system. If the ELT enterprise is to attain its full scientific potential, which is important to us as scientists and to our benefactors, then it must be equipped with the appropriate hardware and, more importantly, software to efficiently add new observational material to the Virtual Observatory where the information will become part of the larger body of multi-facility multi-wavelength data. This topic merits serious discussion, study, specification, and costing in the short term to make sure we do not build-out capabilities that are desirable in the long term.