PROCEEDINGS OF SCIENCE

The Legacy of Discrimination

Elizabeth Griffin[*†]
Herzberg Institute of Astrophysics
E-mail: elizabeth.griffin@nrc.gc.ca

The contents of this paper were presented in three separate sessions of the meeting: *The Legacy of Discrimination*, *Back to the Future* and *Cross-Fertilization and Collaboration*. However, in the context of the meeting they were examined from the aspect of their unifying theme, namely *Data* in astronomy, so for the purpose of this Proceedings they have been woven into a single contribution.

Accelerating the Rate of Astronomical Discovery
August 11-14 2009
Rio de Janeiro, Brazil

[*]Speaker.

[†]I am very grateful to the IAU and to the World Spectra Heritage for financial support to attend this GA

PoS(sps5)018

The motivation for this contribution has been encapsulated by an earlier remark, namely that the traditional share of the budget for a large telescope has been something like 90% for the front end (telescope, dome, etc.), up to 10% for the instruments, and only what is left over for managing the data which the hardware produced. But extracting innovative science from those data fully, and realizing the potential of such equipment, will require a reversal of those figures, with by far the greatest effort going into the data management. Each of my separate views of *Data* describes a different set of challenges which actually result from subconscious bias, the most pervasive being the status of those who used to be the observatory data archivists.

1. THE LEGACY OF DISCRIMINATION

My thesis here asserts that the status of an occupation is substantially influenced by the status of its practitioner. If we delve into the origins of gender stereotyping in astronomy, we can appreciate some of the consequential effects which still colour attitudes towards data archives.

However, re-thinking the situation today is not merely a question of re-casting the budget. Attitudes need to be convincingly re-cast too, and because there are cultural issues at stake it is important to recognize them and to understand their provenance, or the necessary re-cast will not be effective. So we need to look for a moment at the sociological aspects of “data" in astronomy and to appreciate what the re-cast can actually imply.

In world economics, the activities that are taken into account when assessing the prosperity of a nation are those which make money move, absorb resources and produce marketable products. In that context, manufacturing weapons to kill is a “valuable" activity. Yet those who are fully occupied in nurturing a family and caring for the home – tasks which stretch from providing and preparing food, raising and education children, lending help and support to neighbours and generally pursuing peace – are invisible to the system because they are not in paid employment, and what they produce cannot be expressed as marketable items. Given the general tendency for substantially more men than women to be in paid employment, one can appreciate the social roots of the assertion that “what men do is important because it is men who do it", and can understand its converse, “what women do is unimportant because it is [only] women who do it", Since payment for a job represents respect for the person and for the work done and is likely to be proportional to the status of the job in question, the jobs which women have characteristically done in the past within astronomy have become regarded as lowly and insignificant compared to front-line research, largely because it was women who did them.

These simple but inescapable conclusions can explain a great deal. Up to about 50 years ago, “publications" meant hard-copy books and journals, and “data" meant the photographic plates in the plate store, their curators being Librarians and Plate Archivists. Curating those resources was almost invariably assigned to women; the tasks required patience, diligence and thoroughness and a desire to be helpful. Seen from that angle, those tasks exemplified the caring, nurturing roles which women have traditionally occupied, and we can thus appreciate how the post of librarian or plate archivist, and with them the library and the plate archive, came to be regarded as second-rate beside the more prestigious occupations which were mostly dominated by the men at the same institute. There were indeed women in astronomy, as at Harvard; the segregation was vertical, the division

PoS(sps5)018

being the nature of the work. The image of the librarian or plate archivist as a little old lady tucked away in some sheltered nook surrounded by cobwebs was no caricature.

Things changed only relatively recently; Information Technology and computer access for books has made librarianship more respectable, and database technology is also a burgeoning field for human employment. Nevertheless, the stigma left by the pre-computer age still lurks in our subconscious thinking and our expectations. Quite recently a newly-arrived high-profile researcher from the US was discussing the equipping of a new telescope, and I enquired whether there was to be adequate support for a data archive. His response was immediate, like a reflex: “If you had to choose between a new instrument and a data archive, of course you’d choose the new instrument". That person and his peers will become our planning executives within the next decade.

We may have as many new instruments as we want, but they will not accelerate the rate of astronomical discovery unless we have also been able to develop and support adequate, skilful software to manage all the output. The best recipe for truly innovative research will therefore require the re-cast mentioned above, provided it can overcome subconscious as well as conscious bias.

2. BACK TO THE FUTURE

Closely related to this consideration of attitudes towards the curation of data in astronomy is the status of our heritage of (now historic) photographic observations. Astronomy has an estimated 3 million archived plates, distributed in observatory plate stores around the world but with a tendency to cluster at observatories such as Harvard, Mount Wilson, Lick and any European institutions which were productive in the early 20th Century. The rationale for (a) preserving the original observations and (b) making faithful digital versions of the stored information for incorporation into modern research has been argued on numerous occasions. In essence, the availability of data spanning more than one or two decades opens up avenues of research into types of long-term variability which are uniquely important to many areas of astrophysics but which can *only* be studied efficiently if data spanning sufficiently long intervals of time are made accessible.

However, “faithful" digitizing, i.e. capturing all the information recorded in the photographic emulsion accurately and precisely, is not as straightforward as some think. Desktop scanners were designed to copy documents, including snapshots, for appearance and are likely to have positional errors, photometric infelicities and a limited dynamic range, depending on the machine’s type and sophistication and on individual performance, though its cost in relation to a purpose-built digitizer does bring it within reach of smaller observatories and institutions. The desktop scanner certainly has its uses for plates which whose potential is clearly limited (e.g. low resolution, coarse emulsion, restricted depth), and for generating “quick-look" scans of direct plates, but more detailed work usually requires a purpose-built machine to cope with the stringent requirements of the necessary tasks. The PDS scanner was one such machine, produced in the 1970s and 1980s and purchased by many of the larger observatories or institutions. Unfortunately, with the universal transition to born-digital observing the PDS became unwanted, and most were not maintained or even kept; consequently, astronomers wishing to access historic observations nowadays are faced first of all with the challenge of finding where the plates in question can be scanned.

PoS(sps5)018

2.1 Digitizing direct plates

For all its high accuracy and large dynamic range, the PDS is slow. Because of its receding popularity it has also been allowed to become obsolescent. Faced with the challenge of digitizing over half a million large-format direct and objective-prism plates, Harvard Observatory recently commissioned and built a new design of digitizer, the “Digital Access to a Sky Century at Harvard" (DASCH), which can scan an 8×10-inch plate in 1 minute and so could possibly scan the entire collection in 5 years. But Harvard has not had to wait anything like that long to start realizing some of the enormous potential of its (very well maintained) plate collection. For a pilot project, a subset of plates of the “beehive" cluster (M44) was digitized; photometric results quickly identified $\sim$400 large-amplitude variables plus several stars with very mysterious spectra. Since that study involved only about 0.1% of the total collection, we can reasonably expect a great many more such discoveries to emerge.

DASCH is very much a 21st-Century instrument, and it is only by combining historic with modern observations efficiently and reliably with such a machine that we can begin to see results such as these. For Harvard, and equally for the whole of astrophysics, the future promises some extremely interesting science by virtue of creating digital access to certain elements of its past for the very first time.

2.2 Digitizing photographic spectra

At the Dominion Astrophysical Observatory, Canada, we are now able to offer a (limited) scanning service to anyone wishing to digitize plates either from the DAO’s own extensive archive, or ones that are mailed to us. Scanning is carried out on the DAO’s PDS; the products are 1-D spectra in relative intensity units, in equal wavelength steps and normalised to the continuum. The scanning and data reductions are carried out by a trained volunteer technician under the auspices of the *World Spectra Heritage* (a Canadian registered charity), and some payment for his time is in order; a rate of $40 CAN per hour is suggested, but is fully negotiable. How much that will purchase depends on several factors: plate type and size, dispersion, wavelength region, quality, calibration details, vintage, the homogeneity of the sample to be scanned, etc.

The products of the scanning are regarded as the property of the DAO and will ultimately be uploaded onto the website of the *Canadian Astronomical Data Centre* and be available to the astronomical public, but as with any observing activity they can be reserved for the requester’s own use for an agreed period.

One long-term objective is to digitize a major part of the DAO’s own collection (16,000 high-dispersion spectra from the 1.2-m coudé, 93,000 spectra from the 1.8-m cassegrain) using two PDSs, one giving priority to requests and the other scanning large quantities of similar-format plates.

2.3 Other plate-digitizing projects

For nearly 10 years the IAU has supported a Task Force for the Preservation and Digitization of Photographic Plates (PDPP). With a membership approaching 80 or 90, the PDPP acts as a watchdog, and also assists where it can with queries and requests. It publishes a very occasional newsletter, *SCAN-IT*, containing reports of projects on the go or recently completed. Through its pages one can find news about (*inter alia*):

• the APDA (the Astronomical Photographic Data Archive) at PARI (the Pisgah Astronomical Research Institute, NC), which is providing a long-term home for plate collections that are no longer wanted and are in danger of being discarded, and ultimately hopes to digitize many of them,
• scanning the Markarian Survey,
• the Maria Mitchell's collection of 8000 plates, now on-line,
• accessing selected scanned plates from the Carte du Ciel (CdC),
• putting the Potsdam scanned CdC and AC images and catalogue on-line by SkyArchive,
• scanning of astrometric plates with StarScan at the USNO,
• the new digitizer (DAMIAN) at the Royal Observatory Belgium.

There are also worrying appeals from caring souls who see once-treasured plates now languishing in inappropriate conditions, and searching queries regarding the transference of indispensable but often nearly illegible handwritten information (meta-data) that is still in (and only in) observatory log-books.

A census circulated to North American observatories in 2008 counted over 2 million plates in 44 collections, with the 16 largest ones containing 97% of the total. Awareness about the existence and potential of historic photographic observations is definitely on the increase, and the PDPP now has enough international prominence that it is likely to be consulted if an observatory has a decision to make regarding the fate of its collection. However, the community at large yields considerable power in these matters, and it is still a race against time for current and proposed digitizing projects to produce unique science that will convince colleagues of the significant richness of astronomy's preserved data heritage.

3. CROSS-FERTILIZATION and COLLABORATION

My first project to involve the digitization of historic astronomical observations was driven by atmospheric science rather than astronomy, although it incorporated astronomical observations. Every celestial observation made from the ground is obliged to pass through the Earth's atmosphere, and will therefore display any signatures of the atmosphere's constituents which are normally visible at the relevant wavelengths. Figure 1 illustrates absorption by telluric components in the UV; the grave effects they have on ground-based observations are obvious.

All stellar spectra observed in the near ultraviolet show absorption by telluric ozone (O_3) shortward of $\sim \lambda 3400$ Å; Figure 2 illustrates their appearance in the spectrum of Sirius. Named the "Huggins" bands, those features have been monitored in the solar spectrum at Arosa (Switzerland) since 1926, providing us with a unique, continuous dataset of over 80 years' duration (Figure 3). The data in Figure 3 are very "secure" in the sense that each point represents an annual average of something like 300 measurements, but the real variations, caused particularly by local and global weather patterns, make it difficult to interpret accurately the past behaviour of natural ozone concentrations. It is vitally important to get that right: if all of the recent decline is due to anthropogenic interference, then it is within our power to reverse the trends, and investment to those ends will be handsomely repaid. But if (as the dotted green line suggests) there are natural variations as

PoS(sps5)018

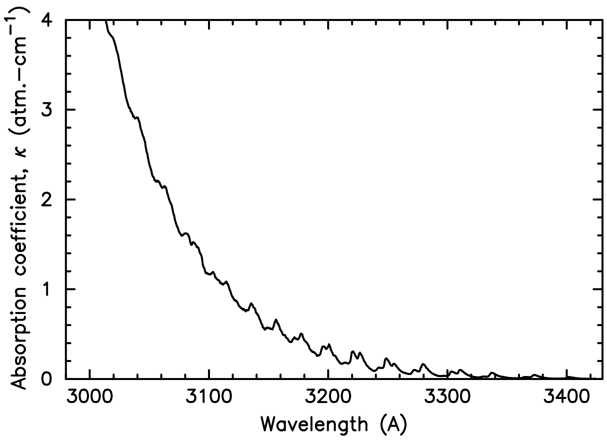

Figure 1: Absorption cross-sections of ozone in the near UV. The intense feature extending shortward is the Hartley band; the small bumps on its redward flank are the Hugging bands.

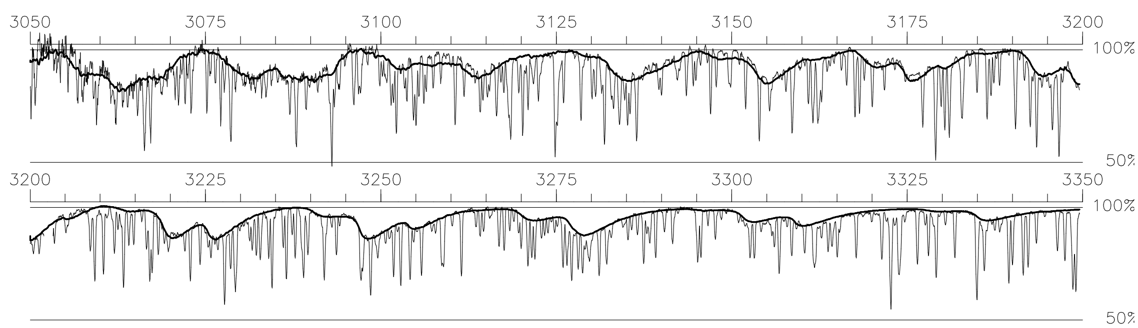

Figure 2: The near-UV spectrum of Sirius. The Huggins O_3 bands are fitted with a laboratory profile.

well, then whatever we invest towards trying to reverse the situation will probably *not* result in the return to pre-1970 values that is intended, and the investment will be seen as a failure.

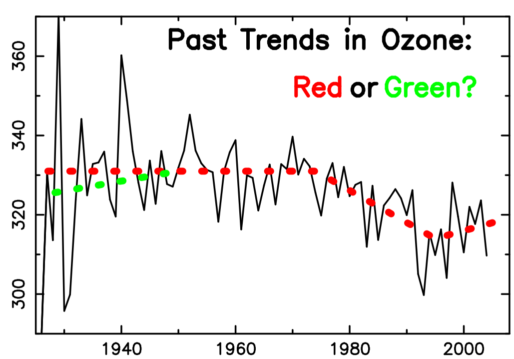

Figure 3: Arosa annual means of O_3 concentrations. The sharp “spikes" depict real variations. It is clear that there has been a steady decline since the 1970s (and perhaps a small recent reversal), but it is less clear whether in earlier decades the mean value was always constant.

Measurements of historic O_3 strengths from independent sources are therefore much needed as supporting data. One such source is our heritage of photographic spectrograms, in particular the spectra of hot stars. But because historic spectra are only available in their native analogue photographic form they have been completely ignored for O_3 work. Moreover, the raw data are not readily comprehensible by a non-astronomer. Thus, while cross-fertilization should have injected some highly sought information from astronomy into atmospheric science, in fact nothing of that sort has happened. When I commenced an investigation of astronomy's photographic plate archives in order to find out the quality and quantity of potential information regarding O_3, I was treading virgin ground.

However, the tasks were less than straightforward on three counts: (a) it required the extraction of an absolute quantity from observations that were only designed to provide information *relative to a star's continuum*, (b) most of the spectra were not optimally exposed for the sort of work which needed to be done and (c) no plate inventories are on-line so visits had to be made in person to each observatory. I therefore had also to build the road along which I needed to travel.

Selecting plates from an archive with no on-line inventory was a somewhat random exercise, but it was soon obvious that only spectra of hot stars would in practice be useful for the kind of direct analysis which I planned to carry out. In addition, the spectrograph optics needed to be of quartz, since ordinary glass absorbs in the near UV, and helpful information on that score could be gleaned in advance from the literature. Papers relating to searches for near-UV interstellar lines also provided a useful guide.

Since the spectra were photographic, they needed to be digitized, and the only operational digitizer of a suitable kind in N America available for the work was at the DAO in western Canada. The borrowed plates were carried to Canada by hand. There was little problem in identifying the Huggins bands, once a template such as Figure 2 was to hand, but measuring their strengths was a challenge because of the band profiles; most features were at least 20 Å wide, so uncertainties in the continuum placement translated into uncomfortably large uncertainties in band strengths. Developing reliable techniques was of course all part of the project. As is customary, my measurements were made relative to the stellar continuum, and in order to convert them into absolute strengths I measured the amount of absorption in the laboratory cross-section curve for each band, yielding a kind of "gf" value. An O_3 strength per plate was thus derived as the mean of however many Huggins bands (mostly between 5 and 10) as could be measured on that exposure, and reduced to the equivalent value at the zenith for that night.

To demonstrate the validity of my method, I selected some high-quality photographic spectra which had been exposed in the last quarter of the 20^{th} Century since an independent source, the *Total Ozone Mapping Satellite (TOMS)* provided its own measurements on the same dates. The essence of this proof-of-concept is illustrated in Figure 4, where the results from 8 digitized stellar spectra are compared to *TOMS* values measured over the same locations; because *TOMS* utilized sunlight and therefore made daytime measurements, I averaged its results for adjacent days to compare with my night-time ones. The fully satisfactory comparison between the values measured by such disparate methods meant that I could then go ahead and obtain O_3 values from photographic spectra exposed on much earlier dates. That work is still continuing.

The project was exciting, but exacting owing to the absence of any precedent for its challenges. A full and thorough working knowledge of photographic spectrophotometry was essential, as was

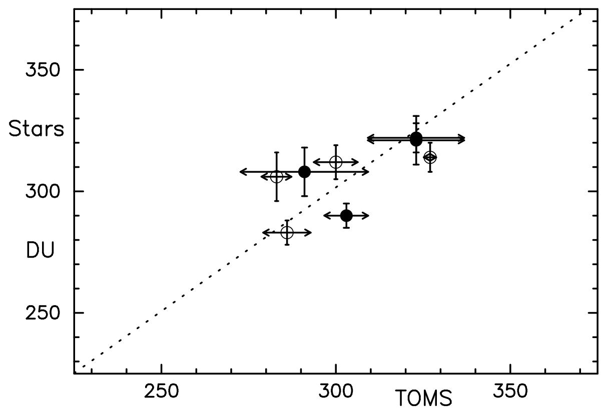

Figure 4: Proving the concept by comparing ground-based and space-based results. Ozone columns (in Dobson units) derived from stellar spectra (ordinates), compared to values measured almost simultaneously by the *TOMS* satellite (abscissae). Filled circles correspond to results from Vega, open symbols to ones derived via Sirius. A Dobson unit is the equivalent thickness of a layer of pure ozone; a characteristic value of 300 DU would occupy a layer 3 mm thick.

also a familiarity with observatory plate archives, so despite its atmospheric-science focus the project was one which only an astronomer could undertake. But the atmospheric scientists secured the support funds, and asked all the right questions (as well as a number of wrong ones). Cross-fertilization was thus very active, but collaboration proved to be more remote, largely because the two disciplines involved did (do) not speak the same language. Nor do our libraries offer access to the literature of the other.

Identifying more such projects will clearly enhance cross-fertilization, but collaboration is something that will only be effective when the data which Astronomy can in principle offer to other disciplines is transformed into a format that can be readily used, and used correctly, outside Astronomy.