

Historical Time-Domain: Data Archives, Processing, and Distribution

Jonathan E. Grindlay¹ and R. Elizabeth Griffin²

¹Harvard Observatory & Center for Astrophysics, Cambridge, MA 02138, USA
email: jgrindlay@cfa.harvard.edu

²NRC Herzberg Institute of Astrophysics, Victoria, BC V9E 2E7, Canada

Abstract. The workshop on Historical Time-Domain Astronomy (TDA) was attended by a near-capacity gathering of ~30 people. From information provided in turn by those present, an up-to-date overview was created of available plate archives, progress in their digitization, the extent of actual processing of those data, and plans for data distribution. Several recommendations were made for prioritising the processing and distribution of historical TDA data.

Keywords. astronomical data bases: catalogs, surveys; STARS: variables; galaxies: active

1. Introduction and Workshop Goals

Historical Time-Domain Astronomy (TDA) data are defined here as significant photographic-plate collections containing exposures of both images and spectra. The importance of those data, not only for current surveys (such as PTF, Pan-STARRS and CRTS) but also for future ones (particularly LSST), was discussed by Grindlay, who emphasized that measurements of the long-term variability of known objects and of historical transients provide long-term comparisons and context—particularly for rare objects or events—for current and future TDA surveys.

The first objective of the Workshop was to gather information about the health and status of astronomy's major plate collections.

The second objective was to ascertain the status of, or plans for, digitizing those collections in order to render the observations fully and freely available and useable. It is only by digitizing the images and spectra that those historic observations are transformed into electronic formats and thus made useful for—and useable by—modern astrophysics investigations and TDA surveys. However, digitization by itself is not adequate; converting the bits is a specialized endeavour and requires specific software that (a) incorporates features which make use of modern all-sky photometric catalogues for both global and local calibration, (b) deals with the point spread function (psf) of photographic rather than CCD images, and (c) involves laboratory spectra for converting stellar spectra into linearized output with the spectrum referenced to an object anchored on the world coordinate system. Plate digitization therefore needs to be accompanied by those additional stages in order to enable accurate scientific analysis, light curves, etc.

The third objective of the Workshop was to survey the state of software for photometric analysis of digitized plate data.

Since these data and analysis tools are only useful if made available in publications and on websites for community access, the fourth objective of the Workshop was to learn about plans and respective time-scales for hosting datasets and analysis tools.

2. Principal Historical TDA Plate Archives and Digitization Status

Seven archived plate collections and their digitization status (objectives 1 and 2) were summarized by their representatives at the Workshop. The archives thus described represented a significant fraction of the world total, which is estimated at 2 million direct-image plates and a further million spectrograms—all in need of specialist digitization for maximum utility in current astrophysics. The respective status of the high-speed scanner in Brussels and of the Wide-Field Plate Database were also discussed.

2.1. *Harvard Observatory, Cambridge, USA*

Grindlay outlined the *DASCH* project (<http://hea-www.harvard.edu/DASCH/>) to digitize the Harvard College Observatory plate collection and make the photometric data available (Grindlay et al. 2009; see p. 29). The total number of direct-image plates is $\sim 450,000$, covering the full sky (with approximately uniform coverage) from ~ 1890 – 1990 except for the “Menzel Gap” between ~ 1954 – 1970 . The custom-designed *DASCH* scanner (Simcoe *et al.* 2006) and its operating software and calibration systems have been further optimized over its ~ 5 -year development, during which $\sim 19,500$ plates have been scanned from 5 selected fields. A semi-automated plate-cleaning machine is in the final stages of development; it will enable cleaning the glass back-side of each plate faster than the 80 or so seconds that it takes for an operating sequence of loading–scanning–unloading a standard-size plate of $30\text{ cm} \times 25\text{ cm}$ (these are scanned two at a time). The “A” series plates are the deepest and have higher resolution ($59''.57\text{ mm}^{-1}$) and are both standard-size (~ 5700 plates) and larger in size ($43\text{ cm} \times 36\text{ cm}$; $\sim 20,000$ plates). The larger plates are scanned singly and require an additional ~ 80 secs per plate to scan. Full “production scanning” and processing (~ 400 standard plates/day) can begin when the two cleaning machines are ready (anticipated for 2012 July). Scanning of the nearly half-million plates together with associated photometric reduction, and the derivation of light curves for all resolved objects, will require ~ 4 years. The $\sim 5,000$ spectrogram plates in the plate collection will be digitized after the direct plates have been completed.

2.2. *DAO/HIA, Victoria, Canada*

Griffin summarized the archive of stellar spectrograms at the Dominion Astrophysical Observatory (DAO). The total of $\sim 110,000$ spectra is the combined output from two telescopes: $>93,000$ from the 1.8-m telescope (inaugurated in 1918) and operated primarily at Cassegrain, and $>16,000$ from the 1.2-m telescope, built with a dedicated coudé in 1962. Since its inception the DAO has pursued the determination of spectroscopic orbits, often for objects with periods of months or years, and while most of the spectrograms exposed for that purpose have probably been measured in the respect of just one variable (the radial velocity), other aspects of the spectra and variable phenomena will have received qualitative assessment only and for the great majority parameters such as temperature, luminosity class and metallicity have never been investigated. Eclipsing variables also feature in the archive, and in particular the chromospheric eclipses exhibited by binaries of the ζ Aurigæ type. Digitization of both sub-sets of the archive has recently commenced using an upgraded in-house PDS scanner, and semi-automated reduction pipelines are now in place for data acquisition, extraction and reduction. Priority is being given to variable objects such as novæ, long-period variables and peculiar stars, but reasonable requests for scanning are also accepted. The fully-reduced data are being mounted on the CADC’s public Website (<http://www.cadc.hia.nrc.gc.ca/cadc/>). The current tally of spectra thus available is approaching 1000 mark, and is steadily rising. The logbooks are also being keyed-in manually. When funds become available, a second PDS will be upgraded and operated in parallel.

2.3. Sofia, Bulgaria

As Tsvetkov explained, while astronomers at Sofia are responsible for the Wide-Field Plate DataBase project, Sofia itself is a repository for the Rozen Observatory (Bulgaria) collection of ~ 9000 plates, plus minor sets from Byurakan Observatory in Armenia, and ESO (La Silla). The combined collection includes direct images from Schmidt telescopes, thereby providing coverage of both the northern and the southern sky; in time they span approximately 1970–1995. Scanning has been started, using Epson scanners.

2.4. Sonnenberg Observatory, Sonneberg, Germany

Tsvetkov described how P. Kroll has digitized (with commercial scanners) $\sim 215,000$ of the total of nearly $\sim 300,000$ plates. Those plates (of the northern sky) cover the period from 1923–1995, and were exposed in various “sky patrol” programmes which were pursued at Sonneberg for decades. Data from Kroll’s digital images are being stored on CDROM as compressed TIFF files.

2.5. Norman Lockyer Observatory, Devon, UK

Goulev summarized the history and status of the collection of ~ 6000 photographic spectra now at Norman Lockyer Observatory in SW England. The Solar Physics Observatory, London, had accumulated “numerous” spectra of bright stars on 8×3 cm and 16×8 cm plates from 1880 to 1912 (when it closed). At that point the plates, *inter alia*, were sent to Cambridge University Observatory (and were largely neglected), but Sir Norman Lockyer, Director of the London site from 1885, continued his work at what was then known as “Hill Observatory” in Devon. Now run largely as an educational facility by local volunteers, the NLO plans to generate a workable database from its own collections of observations, which include a significant number of novæ spectra, in particular of Nova Per (1901). Preservation efforts are under way, and some of the series of plates are already inventoried. When the Cambridge Observatories recently disbanded its own plate store, the collection sent there from London in 1912 was donated to the NLO. Part of any funds raised on the occasion of NLO’s Centenary (February 2012) will contribute towards digitizing the scientific observations.

2.6. Asiago Observatory, Italy

Barbieri described the collection of $\sim 16,000$ plates at the AO, and a current programme for digitization with an Epson scanner. The time required to digitize one plate electronically at 1600×1200 dpi, including cleaning it first, is ~ 20 – 30 min. The digitization programme has been guided by science topic, with an emphasis on AGN variability which has yielded evidence for the binary supermassive black hole in BL Lac itself. The log-books for the plates have also been digitized, and are on-line.

2.7. Ukraine Observatory, Kiev, Ukraine

Pakuliak outlined progress towards putting on-line the photographic observations, totalling some 300,000 direct plates, exposed between 1895–1996 at six Ukrainian Observatories: MAO/NAS, Mykolaiv, Crimean, Kiev, Lviv and Odessa. Each observatory is conducting complementary scientific investigations—from comets to quasars—with its archived material. Resources are limited, but nearly 75% of the plates have been digitized at low resolution using commercial flatbed scanners, and the plan is to do higher-resolution scans based on science priority. Two scans have been made per plate, one rotated through 90° , to quantify and thence minimize instrumental effects.

2.8. *Royal Observatory of Belgium, Brussels*

The ROB was unfortunately not represented at the Workshop. A purpose-built high-speed scanner, code-named *DAMIAN*, has now started scanning, though not solely astronomical material since it is a shared instrument for aerial-photographic films. It has for some time been planned that it serve the broad community by digitizing Europe's astronomical plate archives, but it is still early days. There is also the hope that it will eventually lead to full and public release of data so that, when combined with *DASCH*, astronomy's historical coverage will be enhanced to its maximum potential.

2.9. *Wide-Field Plate DataBase*

A very considerable effort towards realizing the Workshop's first two objectives has been made by the team who have been working very productively on the Wide-Field Plate DataBase (WFPDB), which has its headquarters in Sofia (Bulgaria) under the very capable direction of Milcho Tsvetkov and Katya Tsvetkova. The project started as an IAU initiative under what was then Commission 9, but following the disbandment of IAU Commission 9 *per se* in 2006 it has continued as an independent entity. The development of the Wide-Field Plate Database (<http://www.skyarchive.org>; Tsvetkov 1992), summarised by Tsvetkov, has been accumulating meta-data for $> 2.4 \times 10^6$ wide-field plates in various parts of the world but chiefly in Europe. Details of that project are to be found on p. 417. A synopsis of the current status of the Wide Field Plate Archive (WFPA), which can be accessed at <http://www.wfpa.bas.bg/catalogue.html>, brought the Workshop up to date, and kindled the hope that such comprehensive inventorying be seen through to completion. As Tsvetkov demonstrated, the main Catalogue is a summary table of the characteristics and coverage of 475 astronomical plate collections; another on-line table, the Wide-Field Plate Indexes, contains the parameters of over half a million plates from 133 archives, and can serve up quick-scan (preview) images of some of those; about 250,000 in total have been scanned to date.

3. Photometric Analyses and Data Distribution Plans

The provision of accurate and precise photometry from direct plates, and of spectrophotometry from spectrograms, is essential to reap the full science reward from historic photographic plates. While procedures for coping with the non-linear response of a photographic emulsion to incident light have been further developed since first evaluated by Hurter & Driffeld (1890), one can only approach photometric calibration asymptotically. For direct (imaging) plates, the calibration is best done (in practice, can only be done) by using modern full-sky photometric catalogues to derive spatially- and magnitude-dependent fits to the digital images, as described for *DASCH* in Section 3.1.

For photometric calibration of spectrograms, unless full information in the form of auxiliary exposures to calibration devices is available, approximations are inevitable, particularly where exposures are rather heavy or rather light. The faint end of an exposure also needs careful handling on account of reciprocity failure: an emulsion will fail to record low-level light below a certain threshold, depending on the emulsion type and its state of hypersensitization, if any. Coupled with the provision of calibration marks, such as dots, strips or wedge-readings, is the need for physical measurements that render these informative. Even then, those physical measurements may not provide complete information. If (as not infrequently happened) the source illuminating the raster of holes or strips did not illuminate them evenly, an iteration to correct for that non-uniform illumination is required, though not many observers apparently knew of the need. The matter is non-trivial, since the shape of an emulsion's "characteristic curve" is rarely

represented by enough points to define it unambiguously if some are incorrect through the effect mentioned above.

3.1. *DASCH*

Photometric analyses developed by Laycock *et al.* (2010) and Tang *et al.* (2012) yield rms uncertainties of ~ 0.10 mag over the full range of ~ 100 years of data from the 9 or so series of plates that contribute to a typical light curve, despite differences in plate scale, image quality and systematic effects. The basic approach employs SExtractor as the object detection and isophotal photometry engine for instrumental magnitude determination. Calibration curves are first derived in annular bins to account for vignetting, with instrumental magnitudes fitted against Hubble Guide Star Catalog (GSC2.3) magnitudes (B) for an initial photometric solution. This initial calibration is followed by local corrections to remove spatially-dependent plate effects (usually caused by the emulsion) or sky-related effects (atmospheric extinction, clouds, even trees near the horizon!). The GSC2.3 catalogue is not ideal since its photometric precision is only ± 0.2 mag and it is predominantly in a single band (photographic B). Fortunately, the all-sky APASS CCD survey (www.aavso.org/apass) described by p. 95 has Johnson B and V as well as Sloan g' , r' , i' , which will improve significantly both precision and, particularly, colour corrections for *DASCH* photometry.

The *DASCH* Pipeline (Los *et al.* 2010) and database software run on a high-speed computer cluster and RAID disk system that can process (overnight) the full Pipeline for about 400 plates scanned per day, in production and processing mode, to populate a MySQL database with photometric values and errors for each of the resolved stellar images (typically $\sim 100,000$ on a standard plate; more on A plates). Light curves are generated very rapidly for any object by extracting from the database the magnitudes thus determined from all plates, or only those with magnitude measures meeting a set of user-selected criteria. Variability measures and tests of their validity can then be derived readily. Additional variability analysis tools are being developed, and will be made available when the full database becomes public. To complete each “tile” of sky, all adjacent tiles need to be scanned first to ensure complete coverage; the tile in question can then be released (see p. 29). A “demonstration tile”, 6° centred on the open cluster M44, is available on the *DASCH* Website; light curves and digital images (thumbnails) centred on a given object can be displayed or downloaded for each point in the light curve, and light curves can be readily plotted to either eliminate or include points that have passed a wide variety of Pipeline “filters” (e.g., to remove blended images, plate defects, etc.).

The full *DASCH* output database of $\sim 450,000$ plate images and derived magnitudes for each resolved object (total ~ 1 Pb!) will be made available on disk for world access as it is completed incrementally. The present plans are to digitize the northern sky at Galactic latitudes $|b| > 10^\circ$ first, which enables comparisons first with existing surveys such as SDSS, PTF, Pan-STARRS1 (PS1) and the Catalina Real Time Survey (CRTS). That stage could be completed by mid-2013. Given the difference in limiting magnitudes (< 14 – 18 for *DASCH* against < 20 – 22 for the modern surveys), such comparisons will be mainly for context or extreme transients. The southern sky above/below the Galactic plane will be next, with expected completion in early 2015, followed by the Galactic Plane by mid-2016—or well before LSST. The reason for doing the Galactic Plane and Bulge last is to allow time to develop crowded-field photometry analysis further by invoking point-spread function (psf) and image subtraction techniques in order to improve the present isophotal photometry used for SExtractor. Experiments have recently been undertaken

to optimise the use of PSFEx+SEextractor for magnitude- and position-dependent fitting of the plate psf.

3.2. Asiago

The photometry from scanned images, as presented here by Barbieri, has achieved photometry results with derived uncertainties in the range ~ 0.1 – 0.15 mag (Johnson *et al.* 2005). The Asiago plates are primarily from just two telescope series (see the WFPA website referenced above) and are therefore more homogenous than the much larger number of plate series that contribute to the light-curve data of the *DASCH* output. The psf-fitting routine described in Johnson & Winn (2004) will be investigated for crowded-field photometry for *DASCH*. Unfortunately the limited resources for the Asiago project may preclude full photometric processing, data archiving and distribution.

3.3. Ukraine

The photometric quality of the digital scans of the several plate collections from the six Ukraine observatories is being investigated. However, lack of adequate resources does curtail full-scale processing and data distribution.

4. Prioritization and Future Plans: Message from the Workshop

(1) The benefits of Historical TDA depend critically on the “100-year” dimensions of astronomy’s archived resources. This unique window drives the efforts referred to above and opens new discovery space. However, most of the afore-mentioned projects are seriously short of resources, mostly funding. It seems likely that full processing and public release of historical TDA data will become a reality for the Harvard-/*DASCH* data, but not for other projects. Compared to major current and future surveys, the requirements echoed by this Workshop are small. Investment of even part of the necessary resources will take our science to uncharted depths (e.g., see Tang *et al.*, p. 447).

(2) The Harvard archive suffers from the “Menzel gap” of ~ 15 years (caused by HCO Director Menzel cutting the Harvard plate programme from 1954–1965, but which was only recovered completely by ~ 1970), and priority needs to be given to all data which will help to fill that gap.

Acknowledgements

We gratefully acknowledge support for the *DASCH* programme by the HCO, the NSF (grants AST0407380 and AST0909073), and from the *Cornel and Cynthia K. Sarosdy Fund for DASCH*. We are also grateful to the HIA (NRC, Canada) for temporary funding to continue and enhance volunteer efforts in the plate-scanning department.

References

- Grindlay, J. E., *et al.* 2009, *ASPC*, 410, 101
 Hurter F. & Driffield, V. C. 1890, *J. Soc. Chem. Ind.*, 9, 455
 Johnson, J. A. & Winn, J. 2004, *AJ*, 127, 2344
 Johnson, J. A., *et al.* 2005, *AJ*, 129, 1978
 Laycock, S., *et al.* 2010, *AJ*, 140, 1062
 Los, E., Grindlay, J., Tang, S., Servillat, M., & Laycock, S. 2010, *ASPC*, 442, 269
 Simcoe, R., *et al.* 2006, *Proc. SPIE*, 6312, 17
 Tang, S., *et al.* 2012, *ApJ*, in preparation
 Tsvetkov, M. 1992, *IAU WGWRI Newsletter*, 2, 51