PRESERVING ASTRONOMY'S NORTH AMERICAN HERITAGE RECORDS

Outcome of a Workshop held on 2012 April 18-19 at the American Institute of Physics, College Park, MD, USA

ABSTRACT

Much astronomical archival material is at risk. This material includes many irreproducible observations, principally photographic ones, which together represent a huge source of data regarding the long-term changes of intensity, position and spectral characteristics of astronomical objects. A Workshop dedicated to developing a plan for preserving the archived material – particularly the plates – was held on 2012 April 18 - 19 at the offices of the American Institute of Physics, College Park, MD, USA.

The thirty-nine Workshop participants agreed on the basic features of the Plan. Its prime objective will be to preserve those heritage materials that constitute *observational records* and their metadata on behalf of the astronomical (or wider) community and to have them in a readily accessible and scientifically useful form in order to enable the production of new science. A secondary objective will be to preserve records that are of historical interest. The activities that will be key parts of the Plan are (a) identifying the materials in question, (b) setting protocols and standards for the preservation of these materials and, where necessary, their conversion to electronic form, (c) establishing schemes for systematic cataloging and archiving of the materials, and (d) making the archived data accessible to researchers. Implementation of the Plan will require (a) establishing priorities for action, (b) identifying appropriate collaborating institutions or organizations, (c) promoting awareness of the value of this effort and (d) attracting the requisite funding.

A body is needed to develop the formal Plan, disseminate it to the astronomical community and oversee its implementation. The preference was that this body be a new AAS working group, and an *ad hoc* committee was appointed to prepare a proposal to the AAS for creation of a "Working Group on Time Domain Astronomy" (WGTDA).

INTRODUCTION

The *Workshop on Developing a Plan for Preserving Astronomy's Archival Records* was held on 2012 April 18 – 19 at the offices of the American Institute of Physics, College Park, MD, USA. Co-sponsored by the American Astronomical Society and the American Institute of Physics and with support from the National Science Foundation,[1] the Workshop was organized by the AAS's Working Group on the Preservation of Astronomical Heritage to address the issue that most of the discipline's rich heritage of archival material is presently not readily accessible to researchers and, further, that much is in danger of being lost.

The materials of interest include many valuable observations, principally photographic. These represent a huge, but to date largely unexploited, source of data regarding the changes over time in intensity, position

[1] Award of NSF grant AST-1110231 to the American Astronomical Society, Kevin Marvel PI, is gratefully acknowledged.

and spectral characteristics of astronomical objects. The archival material offers the opportunity to explore behavior in the time domain on scales up to 130 years, many times longer than can be studied with only data from modern digital imaging.

The goal of the workshop was to develop a Plan for, first, ensuring the preservation of material which has scientific value and/or historical interest and, second, having (eventually) the items in electronic and web-accessible form. The Plan would be limited to archival material in North America, but it was recognized that most of what was discussed had relevance worldwide.

Workshop participation was by invitation and was limited to 40 people to ensure that all could participate in the discussions and decision making. Those invited were selected to achieve a good mix of astronomers who use archival data in their research, archivists and historians who deal with astronomical records and people responsible for archiving or digitizing astronomical plate collections. Thirty-nine people attended, coming from twenty USA institutions, two Canadian ones and four located overseas. Thirteen of the fifteen North American institutions having major plate collections – defined as holding over 20,000 plates as reported in the 2008 plate census (Osborn and Robbins 2009) – were represented. A list of participants and their institutional affiliations is given in Appendix A1.

The first day of the Workshop consisted of a series of in-depth round-table discussions framed around a series of "Guiding Questions." The second day began by dividing the participants into small groups tasked with discussing six specific topics deemed critical in developing the Action Plan. This was followed by a general discussion on how to proceed on implementing a Plan. This report provides a summary of the discussions and decisions made at the Workshop.

B. WORKSHOP PARAMETERS

The Workshop first established some basic parameters for the Plan. Three questions were debated and decided:

- What types of materials need consideration in preservation efforts?
- What will be acceptable preservation standards?
- For whom will the Plan be intended?

Materials to be considered

The discussions showed it is most convenient to separate the materials into two categories: (1) photographic plates and their associated metadata (log books, notes on plate envelopes, marking on the plates) and (2) all non-photographic records. The latter includes magnetic tapes, strip-charts, punched tape and cards and paper documents. It was decided the focus will be on materials containing astronomical data. However, besides the observational records themselves, relevant information on calibration and reduction processes (including specialized software) must also be considered.

Excluded from consideration will be documents of purely historical nature, data routinely archived elsewhere (such as, for example, by NASA or the AAVSO), and astronomical instruments (although a few instruments, such as tape readers, necessary for preserving data may need to be part of the Plan). One excluded category for which there is a serious concern, especially by historians, is e-mail which presently is not systematically archived.

Preservation standards

Preservation of the materials obviously is essential. However, not all items have equal potential value. Preservation efforts should be concentrated on what is scientifically important, although that may not be easy to determine.

Having data in digital form is necessary for modern scientific use. Therefore, closely connected to preservation is the issue of digitization – the preparation of digital copies or transformations to electronic form for those materials where they are lacking. In general, the electronic version should be as faithful to the original as can be achieved.

All records will eventually deteriorate, and all can be lost through accidents or disasters. However some formats are more durable than others. With care, photographic plates can last longer than any present digital media, so their preservation should have priority over their digitization. On the other hand, for some materials (such as rapidly deteriorating paper records) the priority may be to secure digital copies as soon as possible.

Plan's audience

The discussions revealed there are four parts to the answer to the question of the purpose and intended audience for the Plan.

- The Plan will serve as a guide to actions needed for the preservation of astronomy's archival records for those responsible for archives.

- The Plan will implement a way to demonstrate the scientific value of this unique resource, with the objective of making the astronomical community aware of its importance and the need to make it more readily available for research. NASA policies require that data from their missions be systematically archived and available to researchers. Should not other astronomical archival data be similarly accessible?

- The Plan will inform those institutions having archives of astronomical records of the expectations and best practices in operating an archive. At present, there is an extremely wide variation. Using plates as an example, some institutions have well-organized collections and a staff archivist, while in other cases the "archive" consists only of boxes of plates stored in a telescope dome, under a stairwell, or in an attic with no one responsible. Some well-established institutional archives have no idea how to deal with astronomical plates.

- The Plan will be used in making a case to potential sources of funding. Archival data preservation and its use will require on-going financial support. The most obvious rational for support is the unique science that can be produced, but projects based on using the material for history, citizen science or data management projects also may be capable of attracting financial support. Scholarly, academic and cultural agencies as well as private sources should all be approached. Commercial organizations were deemed not sufficiently altruistic, nor would they see enough value to them, to be a reliable source of support.

BACKGROUND INFORMATION

After setting the basic parameters for the Plan, background information relevant to its design it was reviewed.

Reference books

(a) ASP Conference Series Volume 410

ASP Conference Series Volume 410, *Preserving Astronomy's Photographic Legacy* (Osborn and Robbins, 2009), contains much useful background information. The book is available through ADS. Relevant article topics include:

- Making the case for preservation of archival material: a white paper on the importance of archival data, the IAU perspective and IAU recommendations on plate preservation, a list of recent research papers that utilized plates.
- A summary of the 2007 PARI workshop on preserving photographic data: the discussion and recommendations, progress made on the recommendations.
- An overview of the Astronomical Photographic Data Archive (APDA) at the Pisgah Astronomical Research Institute (PARI).
- Evaluation of commercial scanners for plate digitization: for direct plates, for spectra.
- Plate digitization projects: at Harvard, at MMO, in Europe, for terrestrial ozone variations.
- Status of North American plate collections: the 2008 census of plates, details on plate collections.

(b) *New Worlds, New Horizons* (the Decadal Survey report)

The findings and recommendations of the 2010 Decadal Survey of Astronomy and Astrophysics are found in *New Worlds, New Horizons in Astronomy and Astrophysics* (National Research Council, 2010). Data archives are discussed in Chapter 5: Sustaining the Core Research Program. While the importance of data archives for modern astronomical research is emphasized, the report's focus is on archiving data acquired in the future rather than on making available observations collected in the past.

Types of archives holding astronomical records

Selected Workshop participants were asked to give brief descriptions of their archives in order to show the range in types of archives and activities.

(a) Harvard's scanning project (*DASCH*) – a mature plate archive (J. Grindlay)

The Harvard plate collection is the world's largest. There is a paid staff archivist. The *Digital Access to a Sky Century @ Harvard* (DASCH) project has been under development for 7 years and has produced the scanner and full analysis pipelines to digitize each plate and produce an astrometric and photometric catalog of all resolved stars on each plate. Some 26,000 plates have so far been scanned in development mode and considerable new science has already emerged. Production processing of the full ~500,000 direct image plates, which cover the full sky from 1885 - 1992, is now beginning and can be accomplished in 4 to 5 years. It will be done in order of Galactic coordinates, starting with the northern Galactic hemisphere (North Galactic Pole down to Galactic latitude +15°), then the southern Galactic hemisphere, and lastly the Galactic Plane (galactic latitudes -15° to +15°). The 10,000 or so plates of spectra, including objective-prism ones, will follow. A total data volume of 1.3 Petabytes is anticipated, of which ~0.6 Pb in direct images and the derived photometry and astrometry will be on disk for direct access. The project's total anticipated cost is $3M. Current funding (from two NSF grants and a Sarosdy Fund gift) will soon be exhausted, and a final significant grant will be required for staff and equipment if the full digitization and data archiving project is to be completed in the desired 3 to 5 years. The *DASCH* website (http://hea-www.harvard.edu/DASCH/) provides a more compete description of this initiative.

(b) Yale Plate Archives – a more typical plate archive focused on science (T. Girard)

The Yale Plate Archives contain a diverse collection of around 100,000 plates, predominantly taken with Yale telescopes or by Yale personnel but including a significant number of plates on loan for measurement. Most records are not electronic; plate storage and organization of the various series are mixed. Nevertheless, many of the plates have high scientific value, and a sizeable number is presently being used for astrometric projects that would be impossible without them. There is a working PDS microdensitometer in a climate-controlled room that is capable of digitizing plates up to 20 x 20-inches; the machine is slow, but has sub-micron positional precision. The archive's focus is on maximizing the science that can be derived from the most important plate series, given the limited manpower and resources available, rather than on (say) collection cataloging. Digitization of the entire collection is not being considered at this time, partly because no digitizer yet exists that is capable of extracting optimally the unique astrometric data on many plates. More details of the Yale Report can be found at http://www.astro.yale.edu/girard/archive/yale.pdf.

(c) APDA at PARI – a repository for others' plates (D. Cline, T. Barker, M. Castelaz)

No plates were ever exposed at PARI, so APDA represents a "specialized archive" – one created specifically for archiving plates that others can no longer keep. A good number of plate collections have been deposited with ADPA for long-term storage, and their log-books are being converted to on-line catalogs. APDA currently holds 200,000 plates and films. The storage buildings have recently been renovated, and climate-control is operational for the plate collection. There is a scanning laboratory that includes several operable machines. The two former STScI Gamma PDS machines are being refurbished and other scanning equipment is being planned. The APDA web page is at http://www.pari.edu/library/apda.

(d) Lowell Observatory Archives – a typical institutional archive (L. Amundsen)

The Lowell archives are representative of many institutional ones in that the holdings are a mixture of archival data (plates) and historical material (correspondence, historical photographs, etc.). All historic log-books back to 1894 are on-line, as also are astronomers' papers, research notes, etc. In addition, a digital museum of 500 scientific instruments is being created. One can browse the holdings at http://www2.lowell.edu/Research/library/; the archives are briefly described by Beiser et al. 2002.

(e) The NRAO Data Archive – an archive specializing in one area (E. Bouton)

Data archiving at NRAO commenced in 2003. The large quantity of NRAO-based papers has now made this archive the *de facto* one for radio astronomy. There are also some digitized audio tapes of interviews with astronomers. Institutional support is good, but funding is limited and the resources are not always sufficient to cope with demands (*e.g.*, the half-inch data tapes had all deteriorated by the time data archiving began in earnest). The web portal is http:\\archive.nrao.edu/archive/e2earchivex.jsp.

Organizations with an interest in archived records

Representatives from several organizations with interests in astronomical archived records summarized what they saw as their organizations' roles in any Plan.

(a) AIP Center of History of Physics and Bohr Library (Greg Good and Joe Anderson)

These AIP centers house collections of papers, books and other items related to physics and astronomy and which are deemed to be at risk and which are of historical importance. They do not support the archiving of data. The primary roles are to encourage the preservation of such materials and to assist researchers in connecting with them and with each other.

Tools for accomplishing their missions include the International Catalog of Sources (ICOS) for locating records worldwide, grants to aid institutions in starting to process archival material (many grants have gone to observatories), and an extensive set of historical photographs and oral histories. The AIP also acts as a matchmaker between those who have papers, collections, etc. and archives; the institution where an individual spent most time is encouraged to be the repository for that person's papers. AIP does accept orphan collections – those that can't find a home but are of significance.

(b) AAS Working Group on the Preservation of Astronomical Heritage (Jim Lattis, Chair)

The WGPAH is relatively recent AAS working group with a charge of "*developing and disseminating procedures, criteria and priorities for identifying, designating, and preserving astronomical structures, instruments, and records so that they will continue to be available for astronomical and historical research, for the teaching of astronomy, and for outreach to the general public.*" WGPAH has no budget and no staff, but the group helps in organizing events such as this Workshop and, when relevant issues are brought to the attention of AAS, the WGPAH is asked to offer advice and suggestions for AAS action. The expertise and concerns of WGPAH are largely historical in nature, and the constitution of the group is not well suited to technical issues involved in acquiring, interpreting, preserving, and digitizing research data. Progress has been made in making the community aware of the need for such records. One step has been reminding astronomers of their obligation to return plates they have borrowed to a suitable archive once they no longer need them.

(c) Virtual Astronomical Observatory (Bob Hanisch, Director)

The VAO is part of an international partnership for providing access to astronomical data worldwide through a single-point portal. Presently there is access to about 10,000 resources. VAO does not archive the actual data, rather organizations keep their data and VAO makes them accessible through data linking using specified protocols for catalogs and data bases. In order for data to be VAO-accessible, data sets must be registered by providing a set of metadata that describes the collection. VAO assists groups with getting their data sets in the VAO format so that they are available for data mining. VAO collaborations include one with Harvard on Dataverse, cloud-like distribution for high-level data products (VOSpace) and science projects with organizations like the AAVSO and others to make data more accessible. The VAO is also collaborating with the Astronomical Source Code Library on access to computer software. Discovery and access is the focus, and they are eager to work with groups with heritage collections of scientific value.

One user's perspective on North American plate collections (Rene Hudec, Czech Republic)

Hudec has been visiting European and North American plate archives seeking plates useful in planning the Gaia mission. From a user's perspective, he has found (a) most North American plates are of high quality, (b) most archives have plates of high value, (c) the available lists of plate archives are incomplete, (d) most places lack reliable funding (often lacking even the most basic things such as proper storage), (e) necessary log books are at times not with the plates, (f) archive holdings are not available electronically, (g) collections are often incomplete (many plates in offices or still on loan), and (h) in some cases access is difficult (the astronomers at some institutions with plate collections are not even aware they have one).

BASIC FEATURES OF THE PLAN

The general features the Plan should have were decided through extensive discussion, some by the group as a whole and others tackled by small breakout groups, as summarized in Appendix A2. It was eventually agreed that:

Objectives

The prime objective of the Plan will be to preserve heritage materials that constitute observational records and their metadata on behalf of the astronomical (or wider) community and to have it in a readily accessible and scientifically useful form in order to enable the production of new science. A secondary objective will be to preserve records that are of historical interest.

Main activities

To realize the objectives, the following activities will be the key parts of the Plan: (a) identifying materials and collections covered by the Plan, (b) setting protocols and standards for the preservation of these materials and, where necessary, their conversion to electronic form, (c) establishing schemes for systematic cataloging and archiving of the relevant materials and (d) making the archived data accessible to researchers.

(a) The materials

The materials to be addressed by the Plan will be mainly the photographic plates and their essential metadata (log books, notes on plate envelopes, markings on the plates). There will be some consideration of non-photographic data records such as magnetic tapes, strip-charts, punched tape and cards and paper documents. Relevant information on calibration and reduction processes (including specialized software) and instrumentation will also be considered when needed for understanding the raw observational records. The Plan will not address astronomical instruments (although ones, such as tape readers, necessary for preserving data, may need to be part of the Plan), documents of purely historical nature, or data routinely archived elsewhere (such as by NASA or the AAVSO).

(b) Preservation and Digitization

Preservation of the materials of interest is essential before the rest of the Plan can be implemented, and preservation will be aided if storage conditions meet archival standards (*e.g.*, those of ANSI, IOA, SAAM). Preservation efforts should focus on what is scientifically important, but this raises the question of how to decide what is "important" given that there may be future uses that we have yet to envision. Another factor, particularly for plates, is that not all items are at their home institutions; missing ones need to be located before they are irretrievably lost and preferably returned so that collections can be as complete as possible.

Having heritage data in digital form is important because modern research so requires. It follows that the plate material must eventually be digitized. Transformation to electronic form should in general be as faithful to the original as can be achieved. Even so, the digitized version is only *an observation of an observation;* no matter how well the transformation to electronic form appears to be, the original plate should be retained as long as is feasible. The number of plates is large but finite, and the cost in time and money needed to digitize all plates could be estimated.

An electronic copy also aids record preservation. Copies significantly lessen the chance that an item will be completely lost, whether through accident, disaster or deterioration. Some record formats are more durable than others – a factor in prioritizing digitization projects. With care, photographic plates will last longer than any present digital media, so their preservation should have priority over digitization. Digitization priorities should emphasize images that offer the greatest scientific potential. On the other hand, for some materials (such as rapidly deteriorating paper records) the priority may be to secure digital copies soon before the records become unreadable. Further points regarding digitization of records are given in Appendix A3.

(c) Discoverability Protocols

Searchable catalogs will be required for efficient access to archived data. Their preparation and use will require both a central database of repositories and standardized systems for classification, organization and listings of holdings.

(d) Accessibility

Modern research not only requires that data to be in digital form, but it also is almost essential that they be accessible through the internet. The Plan must eventually address how this can be done for the heritage data.

Implementation steps

Implementation of the Plan will require (a) establishing priorities for action, (b) identifying appropriate collaborating institutions or organizations, (c) promoting awareness of the value of this effort and (d) attracting the requisite funding. A group will be established to develop the formal Plan and oversee its implementation. This will include dissemination of the plan to observatories, relevant educational and archiving institutions and the general astronomical community with the request that they support its implementation. The Plan and supporting information and will be made available for inclusion in applications for support

(a) Prioritization

Some materials have greater potential scientific value than others, so a scheme is needed for prioritizing both preservation and digitization activities. Factors to be considered include (a) information density (*e.g.*, plates compared to strip-chart recordings), (b) amount of data in the record that have been published (usually close to 100% for strip charts, less than 1% for wide-field plates), (c) format of the data (*e.g.*, radio-astronomy data are already all digital, although early ones are only magnetic tape files and may lack interpretive metadata), (d) condition and rate of deterioration (this affects the urgency to save a record's information, but significantly deteriorated materials – say >80% data loss – may not be worth the effort unless of particular historical value or providing unique information). Setting a preservation criterion based on current scientific grounds is unwise because research interests and needs evolve.

It will also be necessary to establish priorities for the different activities of the Plan as it is implemented.

(b) Collaborating institutions

Venues for archiving data should have an astronomy orientation (*i.e.*, be associated with an institution active in astronomical research). The central database of repositories would preferably be located at a place dedicated to astronomical record archiving, and include suitable mirroring. Materials with mainly historical or cultural content may be archived at centers such as the AIP. Traditional libraries and archives are the preferred repositories for specialized records, such as photographs of people and places or the papers of astronomers and observatories. Consolidated repositories can save cost, but require dedicated personnel.

The AAS has pan-American communication, is strongly supported by the community at all levels of operation, and offers its committees and working groups an established financial structure. It is therefore the preferred host organization for administrating the Plan. A proposal will therefore be made to the AAS to create a "Working Group on Time Domain Astronomy" (WGTDA).

(c) Publicity

Publicity will be a key component for successfully implementing the Plan. Institutions and scientists need to learn the value of heritage material and the attendant responsibilities. To that end, pilot projects which demonstrate the value of archival data to modern astronomical research will be selected and initiated. Projects that focus on outreach, citizen science and history of science will also be encouraged as they can help to attract funding. Increased awareness of the Plan can assist to accomplish some of its targets, such as stimulating the preparation of searchable catalogs and the return of long-borrowed materials.

(d) Funding

The shortage of on-going financial support for archival data preservation and its use is currently a critical problem. Securing funding is essential if the Plan is to accomplish its objectives. Efforts should aim high, but approaches may have different foci. The prime rational for support is the unique science that can be produced, but uses of the material for history, citizen science or data management purposes are also possible. Scholarly, academic and cultural agencies and private sources, targeted for each focus, will be approached. It will be important to understand how receptive agencies such as NSF will be towards these needs. Commercial organizations are deemed not sufficiently altruistic, nor are they likely to see enough value to them in providing support, to warrant much attention.

The cost of operating a plate archive was discussed at the 2007 PARI Workshop (Osborn and Robbins 2009, pg. 57). Major cost areas are utilities, equipment (computers and servers), storage infrastructure (shelving, cabinets), and staff (salary and benefits). This Workshop identified several specific activities presently in need of funding:

- Copying of magnetic tapes that have observations or other data to more readable and permanent media. Data on tapes are probably foremost in danger of being lost as tapes easily degrade and tape-readers are now scarce.
- Physical preservation of plates and their associated documentation (logs, envelope notes, other metadata). Funds are needed to provide and maintain appropriate storage conditions.
- Consolidation of collections. Shipping dispersed plates and other records to their home archives or to another storage location (such as PARI) can be costly.
- Preparation of searchable catalogs of items in the archives. Inventorying and data entry from log books, etc. is labor-intensive.
- Digitization of priority plates and other records. For plates, the expenses include the construction of specialized hardware, software development, and labor.
- Development of on-line (Web) access to archived data.
- Securing of professional expertise for questions or problems relating to archiving, data management, digitization, cataloging, etc.

INITIAL STEPS TOWARD IMPLEMENTATION

It was agreed the first step would be to establish some formal group to oversee on-going work on the issues addressed by the Plan. This group would develop the formal Plan, disseminate it to the astronomical community and oversee its implementation. The strong preference was to request forming a **Working Group for Time Domain Astronomy** (WGTDA) under the auspices of the AAS. An application to the AAS will be prepared by an *ad hoc* Committee. Workshop participants M. Castelaz, J. Glaspey, G. Good, J. Grindlay, E. Griffin, J. Lattis, W. Osborn and I. Shelton were selected as members of that Committee. AAS Executive Director Kevin Marvel will be invited to serve ex-officio, and some representatives of new TDA areas (such as large synoptic surveys) will be invited to participate.

Membership in the proposed WGTDA will be solicited from those with pertinent interests. A WGTDA Executive Committee will be formed to direct the initial activities and set the standards and protocols which the Plan will embrace. The Executive Committee would be composed of about ten members who, between them, represent the following areas of time domain astronomy: data from modern surveys, data from heritage material (including photographic plates), data management, data standardization and integration, follow-up observations of TDA discoveries, and integration of time domain data into virtual observatories. The Committee may add consultants, not necessarily AAS members, as needs arise.

The WGTDA will be charged initially with overseeing the tasks listed below. Groups to work on specific issues, such as creating a master catalog of plate archives, managing data in large surveys, integrating archival data into the VAO, plate digitization, etc. may be established.

Formalize the Plan

The WGTDA under the direction of the Executive Committee will finalize the Plan and prepare a document that describes it and which is suitable for general distribution. It is understood that the Plan may evolve in response to changing conditions.

Disseminate and Publicize the Plan

The WGTDA will distribute the Plan to observatories, relevant educational institutions, archives and professional bodies and the general astronomical community. It will publicize it through announcements and progress reports in the AIP, AAS, HAD and SCAN-IT Newsletters, oral/poster presentations at conferences (including the special session on “Preservation of Astronomical Heritage and Archival Data” at the AAS/HAD meeting in 2013 January). Interest in the preservation initiative may also be stimulated through informational tables or booths at meetings, by engaging students/volunteers in the necessary tasks of record cataloging and digitization and/or by establishing on a WGTDA website a bibliography of current research using archived data.

Select Pilot Projects

Pilot projects will be designed to engage, inform and educate astronomers, historians and the public of the value, potential, and essential nature of heritage material as a unique and important scientific resource for TDA research. To be effective, a pilot project needs to produce outcomes of interest and quality fairly rapidly. Good pilot projects will not only produce solid scientific results but also (a) demonstrate the dependence of some advances on heritage data and (b) open new avenues of exploration for researchers, historians, educators or the public. An objective will be to engender support and interest among varied audiences; one tool may be to tie the use of archived data to a well-publicized new space mission or modern survey. Examples of possible pilot projects are in Appendix A4. They can be grouped into two categories, depending on the audience:

(a) The Scientific Case: New science from old data

Where scientists are the primary audience, a project has to be sold on its scientific merit. The unique feature of heritage data is the long time-base. Some projects can *only* be completed satisfactorily if pre-digital data can be accessed in a scientifically-meaningful form for inclusion in the analysis. A researcher has the responsibility to incorporate all relevant observations in a study, and therefore has a right to expect heritage data to be accessible; the astronomical community has a duty to make them so.

The photographic archives are particularly important. They constitute our only opportunity to extend backwards our investigations of the long-term behavior of celestial objects over time-scales up to 130 years (four times longer than digital imagers have been in regular use, and seven to ten times longer than held by most digital data archives). They also contain important pre-discovery information on objects of new interest. Long time-domain data are fundamental for such studies as the dynamics of the Local Group, binaries with decade-long periods, and phenomena that occur rarely (as exhibited by certain CVs). It should also be mentioned that the technology available when plates were current was not able to extract all of the information latent in even a single photographic exposure. Modern computing capacity and measuring tools now allow considerably more content, detail and complexity to be extracted from plates. Managing this heritage resource responsibly can therefore constitute both a service to present science and a valuable heritage for our successors. However, the window of opportunity to do so is increasingly being limited by the risks of aging and deterioration (especially through poor storage conditions), accidents, and a dwindling knowledge and expertise about photographic plates and their value; the latter has led to valuable collections being discarded. The urgency to act **now** must be emphasized.

(b) Projects Aimed at Other Audiences

Historians and the public also have interests in astronomy's heritage records. Pilot projects focusing on these groups should also be developed. Many of the general public are interested in astronomical advances, and moreover, citizens are entitled to share in the results coming from publicly funded science. Those most interested – at times with a passion – are often attracted by opportunities to be part of the scientific process, and "citizen science" projects have blossomed over the past decade. Some elements of astronomy's archived data lend themselves well to citizen science.

Initiate Access to Heritage Material

Internet access needs to be established early on for selected sets of archived material that have definite scientific potential. Examples could include images from plates that contain a 100+ year record of a specific sky region or digital spectra covering a comparable time period for an enigmatic variable object. A basic first step is to create on-line catalogs of plates or images of log book pages that can be searched for observations of particular importance, but the ultimate goal is to have the actual data accessible through the Web.

References

Committee for a Decadal Survey of Astronomy and Astrophysics. 2010, *New Worlds, New Horizons in Astronomy and Astrophysics* (Washington: National Academies Press).

Osborn, W. and Robbins, L. 2009, *Preserving Astronomy's Photographic Legacy*, ASP Conference Series Volume 410 (San Francisco: Astronomical Society of the Pacific)

Appendix A1 - List of Participants

NAME	AFFILIATION
Amundson, Lauren	Lowell Observatory Library, Flagstaff AZ
Anderson, Joe	AIP – Niels Bohr Archives, College Park MD
Barker, Thurburn	Pisgah Astronomical Research Institute, Rosman NC (Director of APDA)
Bartlett, Jennifer	U.S. Naval Observatory, Washington DC
Blanton-Kent, Beth	University of Virginia Library, Charlottesville VA
Bouton, Ellen	National Radio Astronomy Observatory, Charlottesville VA (NRAO)
Castelaz, Mike	Pisgah Astronomical Research Institute, Rosman NC
Cline, Don	Pisgah Astronomical Research Institute, Rosman NC (PARI President)
De Cuyper, Jean-Pierre	Royal Observatory of Belgium, Brussels, BELGIUM
DeVorkin, David	National Air and Space Museum, Washington DC
Gates, Ellie	Lick Observatory, Mt. Hamilton CA
Girard, Terry	Yale Observatory, New Haven CT
Glaspey, John	NOAO – Kitt Peak Observatory, Tucson AZ
Good, Greg	AIP – Center History of Physics, College Park MD
Griffin, Elizabeth	Dominion Astrophysical Observ., Victoria BC, CANADA
Grindlay, Josh	Harvard College Observatory, Cambridge MA (DASCH Director)
Hanisch, Robert	Space Telescope Science Institute/VAO, Baltimore MD
Hartkopf, Bill	U.S. Naval Observatory, Washington DC
Hudec, Lukas	Czech Technical University, Prague, CZECH REPUBLIC
Hudec, Rene	Academy of Sciences of Czech Republic, Ondrejov, CZECH REPUBLIC
Jacob, Andrew	Sydney Observatory, Sydney, AUSTRALIA
Kraushaar, Andy	Wisconsin Historical Society, Madison WI
Lattis, Jim	University of Wisconsin-Madison, Madison WI (WGPAH chair)
MacConnell, Jack	Towson University, Towson MD
Marvel, Kevin	American Astronomical Society, Washington DC (AAS Executive Officer)
Mills, O. Frank	Yerkes Observatory, Williams Bay WI
Mills, Jill	Yerkes Observatory, Williams Bay WI
Mueller, Jean	Palomar Observatory, Palomar Mountain CA
Osborn, Wayne	Yerkes Observatory, Williams Bay WI
Pascu, Dan	U.S. Naval Observatory, Washington DC
Patterson, Ricky	Univ. of Virginia – McCormick Observatory, Charlottesville VA
Rees, Rick	Westfield State University, Westfield MA
Rottler, Lee	Pisgah Astronomical Research Institute, Rosman NC
Shelton, Ian	University of Toronto/Mt. Alison University, Toronto ON, CANADA
Simcoe, Bob	Harvard College Observatory, Cambridge MA
Sliski, David	Harvard College Observatory, Cambridge MA
Stein, John	Allegheny Observatory, Pittsburgh PA
Tsvetkov, Milcho	Bulgarian Academy of Sciences, Sofia, BULGARIA
Tsvetkova, Katya	Bulgarian Academy of Sciences, Sofia, BULGARIA

Appendix A2 – Discussion Topics

General Discussions

1. **The Plan's Objectives:** Why is a Plan needed, and what should the Plan be designed to achieve?
2. **Breadth of the Plan:** It is not feasible to develop one Plan that would be applicable for all types of archived material. Separate plans for the various types of records will be needed, so how broad will the focus be for this Plan? What materials will be included, and what features will this Plan need to ensure that it will be a useful prototype for future plans for other materials?
3. **The Issues:** What issues must the Plan address?
4. **Key Parties:** The desirability of specialized archive or digitization centers and the centralization of activities will be a key factor in designing the Plan. Equally important will be the parties that must provide the necessary resources (*e.g.*, funding) for implementing it. To what extent should the Plan be based on specialized centers, what should be the financial expectations of those involved in, or benefiting from, archiving, and could funding be pursued through agreements with commercial archiving or database firms?

Small Group Discussions

Six topics were selected for discussion by small groups or five to six participants, selected so as to contain a mix of astronomers and other professionals. Each group discussed two of the topics, so each topic was independently treated twice. The results from the discussions were later summarized in a plenary session and the views from the two separate discussions compared.

1. **Science from archived data**: The Plan should include a statement of the kind of science that can be carried out with archived data so as to awaken and excite interest. What are some possible examples?
2. **Community cooperation and structure:** How do we maintain the momentum? How can this workshop transform itself into an organizational structure for dealing with archived data in North America? What special interest groups need to be included, such as based on technical expertise or on the nature of certain materials? A vibrant and sustainable mission is needed.
3. **Making archived data accessible to scientists, other academics, and the public:** How do we provide web access to digital information? How do we make the project known and make the archived data VO compatible? Some specific questions are what form do the data take, what does the user need (including metadata), is there more to this than a website, how is compatibility achieved (VO compatibility, compatibility between archives)?
4. **Funding**: What are the priorities for funding? How might funding be applied?
5. **Effective pilot projects:** What projects will best engage research astronomers, historians, the public, and others?
6. **Photographic plate archives:** What are appropriate criteria and standards for preserving the information in photographic plate archives including the associated metadata (such as in log books). What are the selection criteria for information specific to astronomers, to historians and to the public and for use in citizen science projects?

Appendix A3 – Comments regarding Digitization

1. There is a tendency to use such terms as “digitizing” and “scanning” interchangeably. For the Plan,
   a. **Digitization** is the creation of an electronic record that is a *faithful copy of the original record* or its information content and suitable for high-level analysis (examples: high resolution versions of direct plates, fully-reduced digital spectra from photographic images).
   b. **Scanning** is creation of an *image of a record* by a technique that is inferior to digitization but is suitable for visual purposes (*e.g.*: thumbnail images of plates) or low-level analysis.
   c. **Transcription** is the *conversion of written records into something that can be searched electronically* (example: creating a database from log book pages).

2. The ideal is to digitize each record only once using as high a quality method as possible. However, digitization standards for data records, such as plates, are typically more stringent than for documents. Telescopes and their instruments inherently impose their own instrumental signatures (*e.g.*, wavelength region, resolution) upon an observation, and anything that further modifies the observational data should be avoided (for example, use of an unsuitable scanner to digitize a plate will impair the quality of the digital copy and reduce its scientific applications). Digitization standards may be relaxed for documents and for some citizen science projects, and a sub-optimal digital copy may be better than a badly deteriorated original.

3. Digitization of the photographic records is crucial, but has inherent problems.
   a. While it would be best to digitize each plate once, many plates contain more information than current digitizers can extract. Plates that have been digitized must therefore be preserved, both as back-up and for possible further analysis.
   b. A purpose-built digitizer may sacrifice one parameter (*e.g.*, photometry) to enhance another (*e.g.*, digitization speed). Any such output limitations should be made clear to users of the output.
   c. Centers for plate digitization could be established. Systematic operation at high fidelity would ensure homogeneous quality of output. It is debatable whether shipping plates to a center with suitable capability is preferable to building and operating traveling digitizers. Relative cost is the major factor but is difficult to estimate at all accurately. It would probably be more expensive to build enough duplicate machines for a single center to achieve a reasonable throughput than to build a few mobile units (such as several high-quality digitizing heads that could be installed on bases supplied locally). There are other factors: the operation of digitizers requires trained staff, multiple centers increases overhead, shipping involves costs of labor (packing/unpacking) and transport and increases risks of damage.

4. Some benchmarks from the Harvard *DASCH* project:
   a. In production mode *DASCH* expects to process 300 – 400 8 x 10-inch plates per day, or ~100,000 per year (the digitizer can accommodate smaller and somewhat larger plates as well).
   b. Full-time production requires ≥ 6 trained staff, including those for plate-handling,
   c. Cloning the *DASCH* Digitizer would cost ~ $250,000, but less for only the imaging head.

Appendix A4 – Projects Using Archived Data

There are many cases where archived observations have played an important – often decisive – role. For instance, HST positioning depends on data from photographic plates. Listed below are research areas where archived observations are important and some possible effective pilot projects.

Research areas needing archived data

1. <u>Near earth asteroid studies</u>: refining the orbits of near-earth objects and discovering new potential ones. Historic observations of objects found (for example) in the Near Earth Asteroid Tracking (NEAT) or the Palomar Transient Factory surveys can be critical for assessing hazards.
2. <u>Supernova research</u>: recovery of supernova precursors, time-studies of progenitors (especially Type Ia) and observational and statistical studies of rare classes of outbursts.
3. <u>Stellar astronomy in the time domain:</u> discovery and study of new classes of objects and investigations of rare phenomena. *DASCH* has discovered new types of stellar variability (including effects attributed to stellar-mass black holes) and mass-transfer episodes without the expected nova events. Rare phenomena include some CV behavior and eclipsing binaries, such as ε Aur, with decade-long periods.
4. <u>Galactic structure and dynamics</u>: proper motion studies related to galactic interactions, dark-matter halos, cluster disruptions by molecular clouds and Local Group dynamics.
5. <u>Historical research</u>: papers, correspondence, log books, plate notations, etc. of astronomers, referees of papers and directors of observatories. Synthesizing documentation from different sources can lead to new interpretations
6. <u>Trans-disciplinary studies</u>: historic record of concentrations of stratospheric ozone; quantifying changes in light pollution.

Possible Pilot Projects

<u>To engage astronomers</u>:

The most appropriate pilot project for an astronomical audience was deemed to be conducting a systematic survey of a broad region of the sky, employing a long time-base and high cadence. *Rationale*: integrating long time-base observations with modern survey data will lead to the discovery and study of new phenomena (transients, anomalous events) and will improve statistics for rare events. This project could be focused on supporting *DASCH* by providing additional data for the new classes of objects being discovered, or by filling in the "Menzel gap" in the Harvard data.

Other possible projects include (1) time-domain spectroscopic studies of long-period variability, (2) following-up interesting objects detected by modern large surveys, in the Kepler field, etc. using heritage data and (3) utilizing archival data in support of space missions (*e.g.*, Gaia; New Horizons).

<u>To engage historians</u>:

A suitable pilot project needs more input from the historian community. One suggestion was a comprehensive study of historical photographs, which might include astronomical discovery plates, and their related documentation. *Rationale*: Modern analyses of plates can reveal what was missed, while putting discovery material into a modern context may shed new light on what led to the discovery.

<u>To engage the public</u>

Effective citizen science projects could include the review of digitized images of plates for (a) asteroid identifications, which would include finding pre-discovery observations and previously missed NEO's, (b) light curve determination of long-term variables and (c) optical detections of gamma-ray bursters and other transient phenomena. *Rationale*: The opportunity to make a discovery will attract citizen scientists, while the multiple-purpose use of the material will demonstrate its high information content.