

The Harvard Plate Stack Scanning Project: 100 years of Legacy Data

Doug Mink	Software and archive	<i>Smithsonian Astrophysical Observatory</i>
Alison Doane	Plate Curator	<i>Harvard College Observatory</i>
Bob Simcoe	Digitizer Design	<i>Harvard College Observatory affiliate</i>
Ed Los	Digitizer Interface	<i>Harvard College Observatory affiliate</i>
Josh Grindlay	Principal Investigator	<i>Harvard University Astronomy Department</i>

Presented at the Building Observatory Legacy Archive BoF at ADASS XVI, October 16, 2006

- **What We have**

500,000 astronomical plates

- **What We've Done**

Imaged logbooks

Transcribed catalogs

Built a digitizer

Digitized >500 plates so far

- **What We're Doing**

3-year NSF grant to digitize 10,000 plates

- **What We Want To Do**

Digitize them all and serve them on the Web

What We Have

- Harvard's plates contain the most complete sky coverage of both the northern and southern sky over the longest time period – 1880 to 1989
- Harvard Observatory now has on the order of 500,000 photographs, by far the largest collection and 25% of the world's total.
- Of those, between 250,000 and 350,000 will be useful for photometry and astrometry.

What We Have

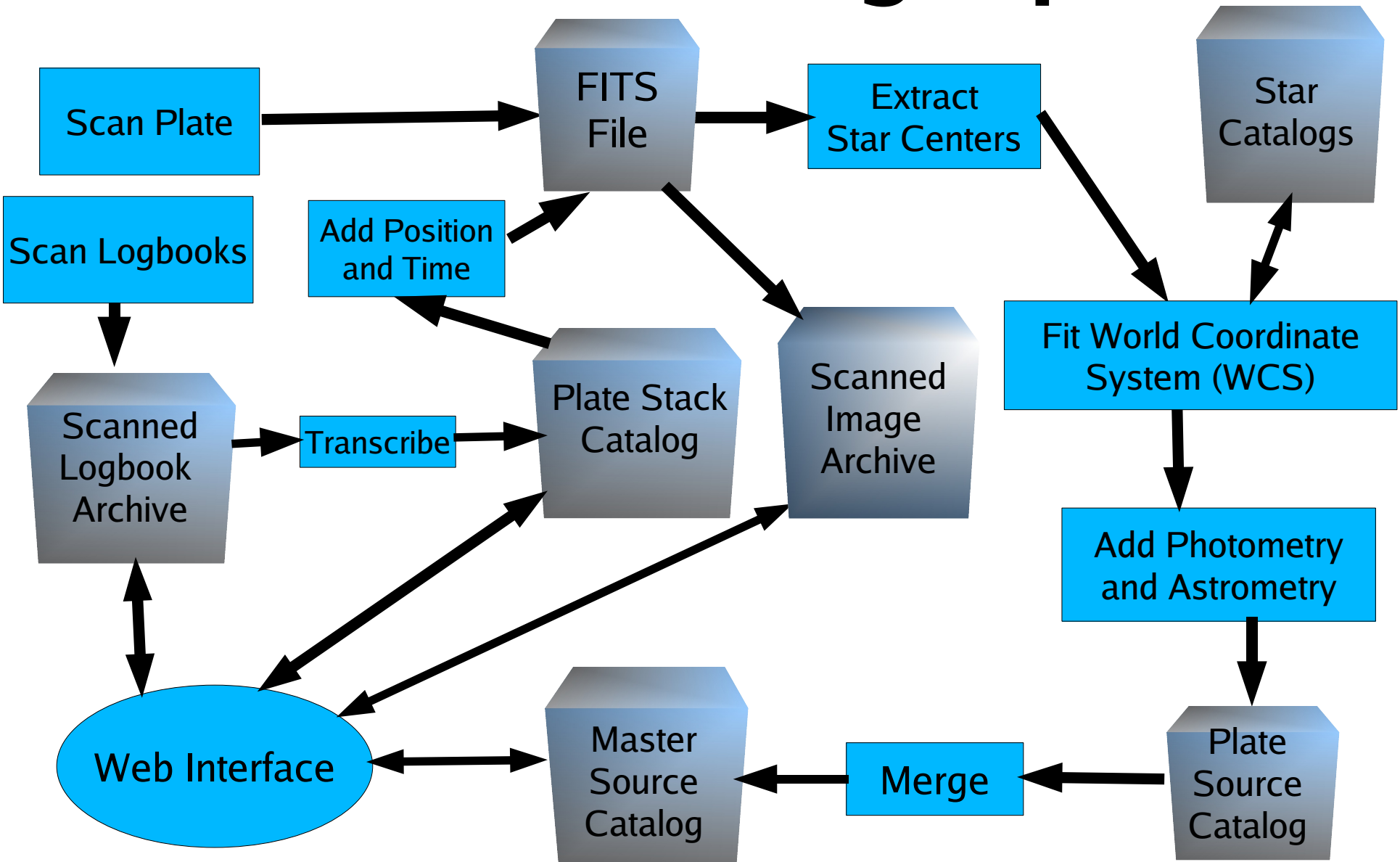
These are the plates which we think are worth scanning

Series	Total	Aperture (in.)	Scale "/mm	N/S	Years	Mag. Limit	Q
A	27504	24	60	S	1893-1950	18	5
ADH	7067	30	68	S	1950-1963	18-19	4
AM,AC, ...	75000	1.5	600	N/S	1898-1957	13-14	3
B	76874	8	179	S	1885-1954	17	4
BR	4176	8	209	S	1938-1944	17	4
DNB,DSB	9000	1.6	580	N/S	1962-1989	15	5
I	59246	8	163	N	1889-1946	17	4
MA	11737	12	97	N	1905-1983	17-18	5
MC	40596	16	98	N	1909-1992	17-18	5
MF	40897	10	167	S	1915-1955	17	4
RH,RB	33000	3	391	N/S	1928-1963	15	3
Total	385097						

What We Have Done

- Created digital catalogs of about 140,000 plates
- Started imaging observing logs, page by page
- Built a digitizer and digitized 545 plates for a proof-of-concept photometric project (as of 2006-10-25)
- Created a web site with access to plate stack catalogs, observing logs, and digitized plates

Plate Scanning Pipeline

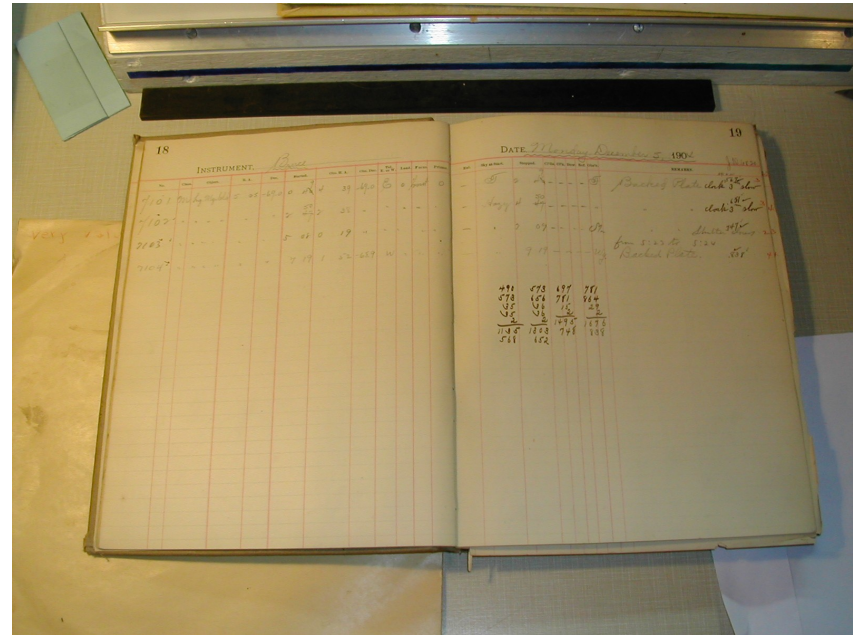


Before We Scan the Plates

- To process plate scans, we need to have digital catalogs. We have a 140,000-plate head start, but the rest of the plate stacks need to be digitally catalogued in a more timely way.
- To outsource catalog transcription, we need to convert existing observing log books to digital images, a fairly major project

First: Digitize Metadata

From hand-written cards and logbooks



Telescope Log Digitization

- Since many of the older log books are too fragile to travel and transcription in Cambridge is likely to be too expensive, they are being imaged into a digital format
- Log books contain information which may not fit or have been processed into existing digital catalogs.
- By the end of 2006, a volunteer will have imaged half of the log books.
- The first 40 volumes have been transcribed in India and are being proofread before conversion into our catalog format.

Catalog Digitization

- Existing digital catalogs have been created very slowly over the past 15 years by part-time plate stack staff.
- The first 40 volumes of image logbooks have been transcribed in India and are being proofread before conversion into our catalog format.
- We are applying for grants to fund further catalog transcription.

Next: Digital access to image data

Moving the plates out of the 20th century



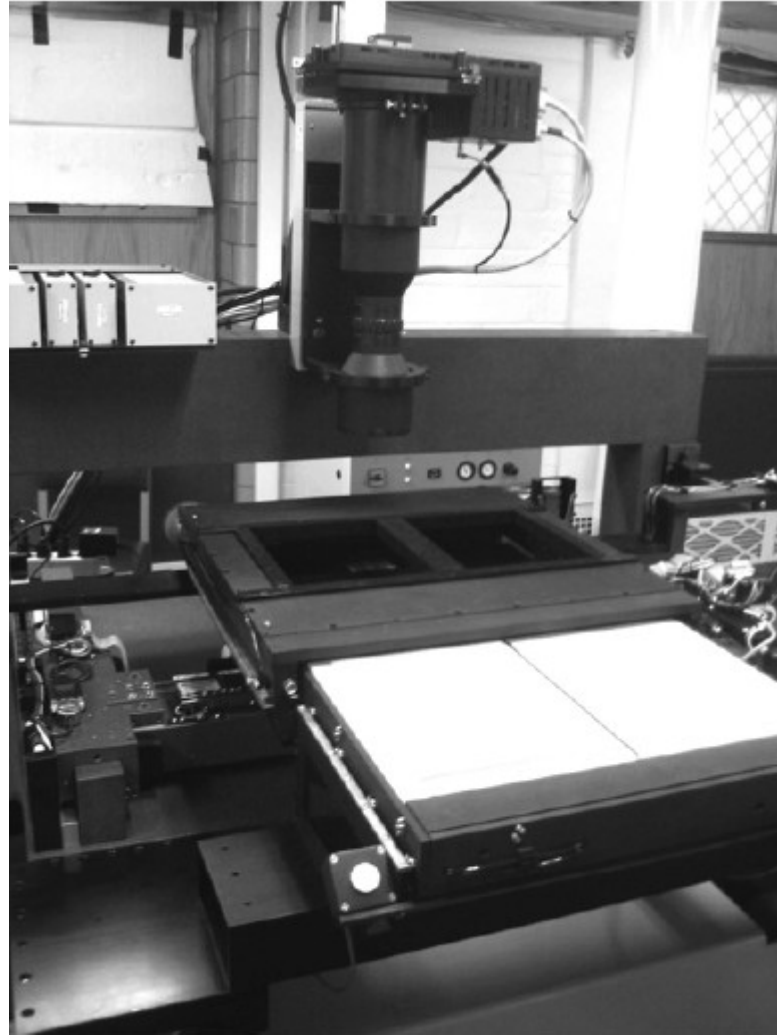
Image Digitization Issues

- To scan Harvard's library of historic plates in a 3-5 year time frame, we needed a machine that can scan 200 times faster than machines designed 20+ years ago, such as the USNO PMM, which took ~1- 4 hours to scan a single 14 x 14 inch plate.
- To meet astrometric, photometric, and archival goals, the machine needs sub-micron positional accuracy, at least 12 bits of photometric density range, and a scan speed that allows human handling to limit the average plate processing time.

Image Digitization Solution

- Using technology common to semiconductor wafer and flat panel display inspection stations, we built a machine that can do ultra fast, ultra precise scanning.
- Plates are moved under a 4Kx4K digitizing camera with 16-bit 11- μ m pixels, taking overlapping exposures.
- An 8 x 10 inch plate is scanned in about 40 seconds generating a 740-megabyte image, though we digitize 2 at once in under a minute and a half.
- A 14 x 17 inch plate can be scanned in a little over a minute, generating a 2.3 gigabyte image.

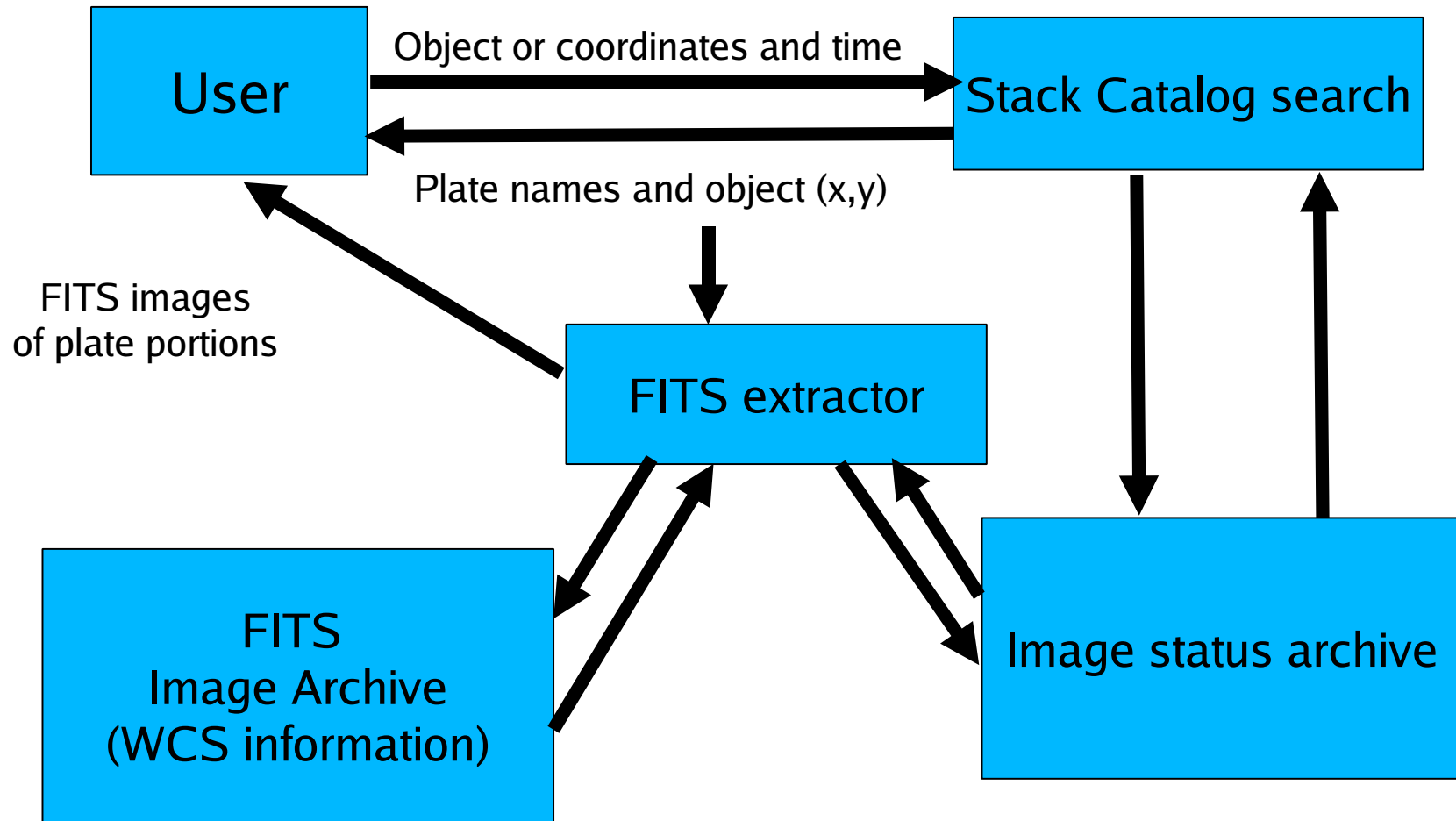
Harvard Digitizer with two 8x10-inch photographic plates



Putting It All Online

- A page of metadata for each plate series describes how and where the plates were taken and points to graphical descriptions of coverage in space and time.
- A search page for each series shows spatial coverage and allows users to find plates covering a specific location in space and time.
- On the list of plates available, those which have already been scanned can be viewed in what we call “elephant nail” full screen images which are reduced by a factor of 36 from the full scans.
- Portions of the viewed plate may be downloaded as FITS or JPEG images. Eventually this will be a VO image cut-out server.

Access to digital images



Access to digital images

File Edit View Go Bookmarks Tools Help
http://tdc-www.harvard.edu/plates/mcseries.html
Internet Lookup New&Cool Google AstroResources MassBike SAO TDC Weather Harvard Plates

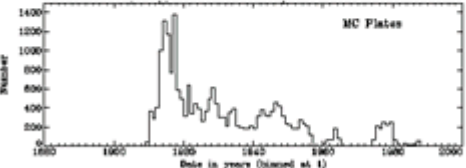
Telescope Data Center
Harvard-Smithsonian Center for Astrophysics
Harvard Observatory Plate Stacks
MC Series Characteristics

Basics
Plate Size Plate Scale Telescope Focal Length
8x10, 4x5 987/mm 16-inch Metcalf Doublet 83 inches

Observatory
Plate nos. Location Dates
1-26093 Cambridge 1909-06-22 - 1932-05-28
26094-40591 Oak Ridge 1932-09-28 - 1988-12-12

[Online Searchable catalog](#)

[Maps of spatial and temporal distribution](#)



[Histogram by year](#)

ASCH Catalog

- [MC Series tab-separated table \(number-sorted B1900 Starbase MJD\)](#)
- [MC Series tab-separated table \(number-sorted B1900 Starbase ISO date\)](#)
- [MC Series tab-separated table \(number-sorted J2000 Starbase ISO date\)](#)
- [MC Series tab-separated table \(RA-sorted B1900 Starbase MJD\)](#)
- [MC Series tab-separated table \(RA-sorted B1900 Starbase ISO date\)](#)
- [MC Series tab-separated table \(RA-sorted J2000 Starbase ISO date\)](#)

Location in Stacks
Plate Size Location Numbers
8x10 first floor all
4x5 first floor all

Card Catalog: complete; there is also an Approach database of the blue plates.

Record Books: (49) 1-37502 (Catalog)
(2) MC Important Regions
(2) MC Long Period Variables
(1) MC Photometric Work

Record Sheets: 26154-40591

Comments: This instrument, the most powerful refractor in the northern station up to 1936, was especially valuable for faint objects where the large scale was of benefit. When the plates are used for photometric purposes, it is well to take precautions, since the color coefficient is large and the distance correction is both large and variable from one plate to another.

The instrument was refigured after about 3500 plates had been taken; the distance correction was thereby diminished, but the color coefficient was practically unchanged. In general, it is well to avoid plates taken before the refiguring for accurate photometric work. All plates subsequent to 4171 were taken with the curved plate holder, which diminishes the distance correction due to curvature of field.

Done

MC Series description page with spatial and temporal distribution

Access to digital images

File Edit View Go Bookmarks Tools Help

http://tdc-www.harvard.edu/cgi-bin/plates/mcsearch

Internet Lookup New&Cool Google AstroResources MassBike SAO TDC Weather Harvard Plates

Telescope Data Center **Harvard Plate Stacks** [A Series \(about\)](#)
[MC Series \(about\)](#)
[MF Series \(about\)](#)
MC Series Search

Name 5-digit Sequence or Object Name X pixels to extract Y pixels to extract

Right Ascension (hh:mm:ss.sss) Declination (dd:mm:ss.sss) Extracted Image Type

Starting Date (yyyy-mm-dd or frac. year) Ending Date (yyyy-mm-dd or frac. year)

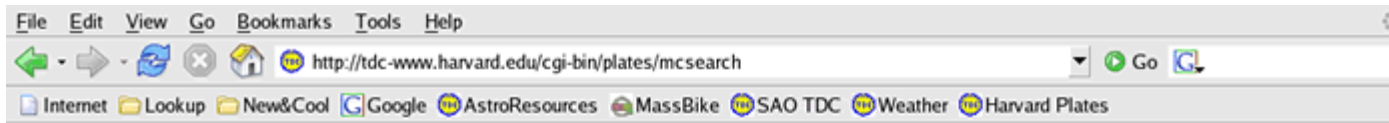
Click on map to see MC plates from 1909-1919


· Harvard MC plate stack catalog (mobra)

Search page for MC series

Limits may be set in space and time

Access to digital images




Harvard Plate Stacks
[A Series \(about\)](#)
[MC Series \(about\)](#)
[MF Series \(about\)](#)
MC Series Search

Name: X pixels to extract: Y pixels to extract:
 J2000 Right Ascension (hh:mm:ss.sss): Declination (dd:mm:ss.sss): Extracted Image Type:
 Search Starting Date (yyyy-mm-dd or frac. year): Ending Date (yyyy-mm-dd or frac. year):

Searching for plates containing m44 = 08:40:22.198 +19:40:19.43 J2000 from NED from 1950-01-01 to 1980-01-01

Plate	RA2000	Dec2000	Exp	Epoch	Arcsec	Comment
38526	08:50:51.624	+19:48:44.96	10.50	1961-05-04	8900.71	Praesepe_Grating_#
38527	08:40:52.321	+19:49:17.87	10.00	1961-05-04	686.14	Praesepe_Grating_#
38529	08:40:52.321	+19:49:17.87	12.00	1961-05-05	686.14	M44_Grating_#2301
38531	08:40:52.321	+19:49:17.87	8.00	1961-05-18	686.14	Praesepe_M44
38531	08:32:52.854	+19:49:45.09	8.00	1961-05-18	6368.78	M44_Grating_#2301
38532	08:40:52.321	+19:49:17.87	1.50	1961-05-18	686.14	M44_Grating_#2301
39048	08:40:52.321	+19:49:17.87	7.00	1964-04-12	686.14	Praesepe_NGC_2632 scanned plate
39053	08:40:52.321	+19:49:17.87	5.00	1964-04-17	686.14	Praesepe_NGC_2032
39237	08:38:52.338	+19:42:24.61	15.00	1975-01-18	1275.25	M44_clear scanned plate
39715	08:40:51.000	+18:30:17.91	22.00	1977-05-13	4221.31	Praesepe_no_filter scanned plate
39927	08:50:49.699	+17:48:45.00	30.00	1978-02-28	11146.90	Patrol_no_filter scanned plate
39927	08:50:49.874	+17:59:45.00	30.00	1978-02-28	10761.59	No_filter scanned plate
39951	08:38:52.644	+20:00:24.60	30.00	1978-03-10	1746.16	M44_no_filter
39963	08:40:52.203	+19:42:17.88	29.00	1978-04-10	440.01	M44_no_filter scanned plate

Results of search for MC plates from 1950-1989 on which M44 appears

Plates which have been scanned may be displayed

Done

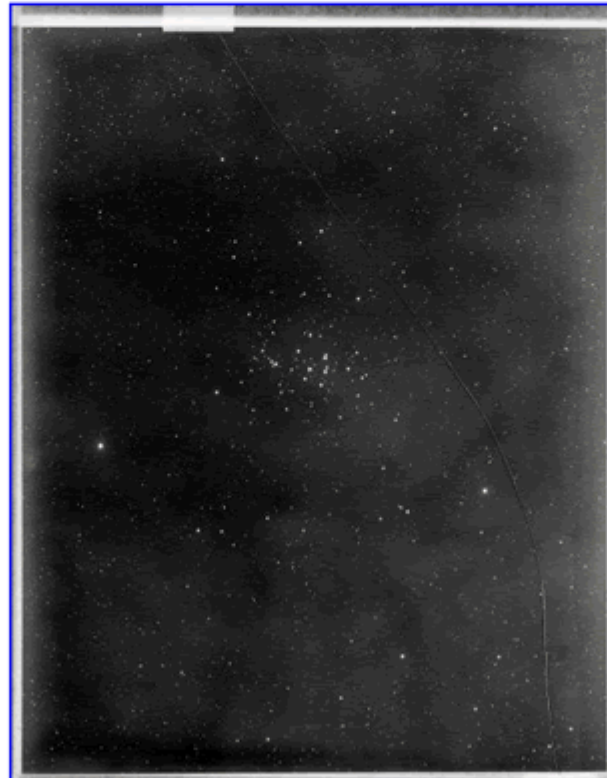
Access to digital images



 **Telescope
Data Center**
SMITHSONIAN ASTRONOMICAL OBSERVATORY
[Click here for 1/20th scale full FITS image](#)
Click on image for x FITS image.

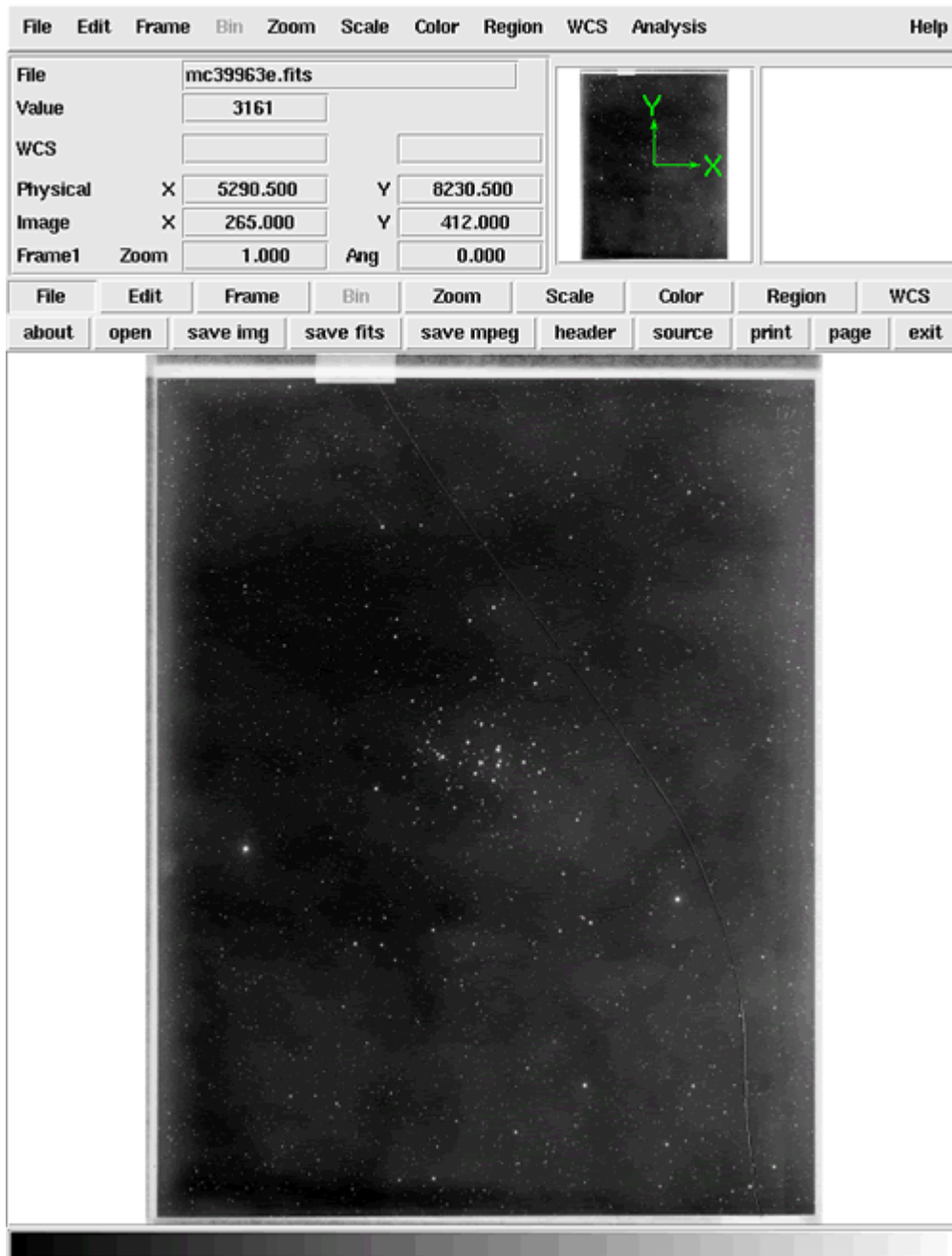
Harvard Plate Stacks [A Series \(about\)](#)
[MC Series \(about\)](#)
[MF Series \(about\)](#)
MC Series Plate
MC39963

Plate	RA2000	Dec2000	Exp	Epoch	Note
39963	08:40:52.220	+19:43:17.88	29.00	1978-04-10	M44_no_filter



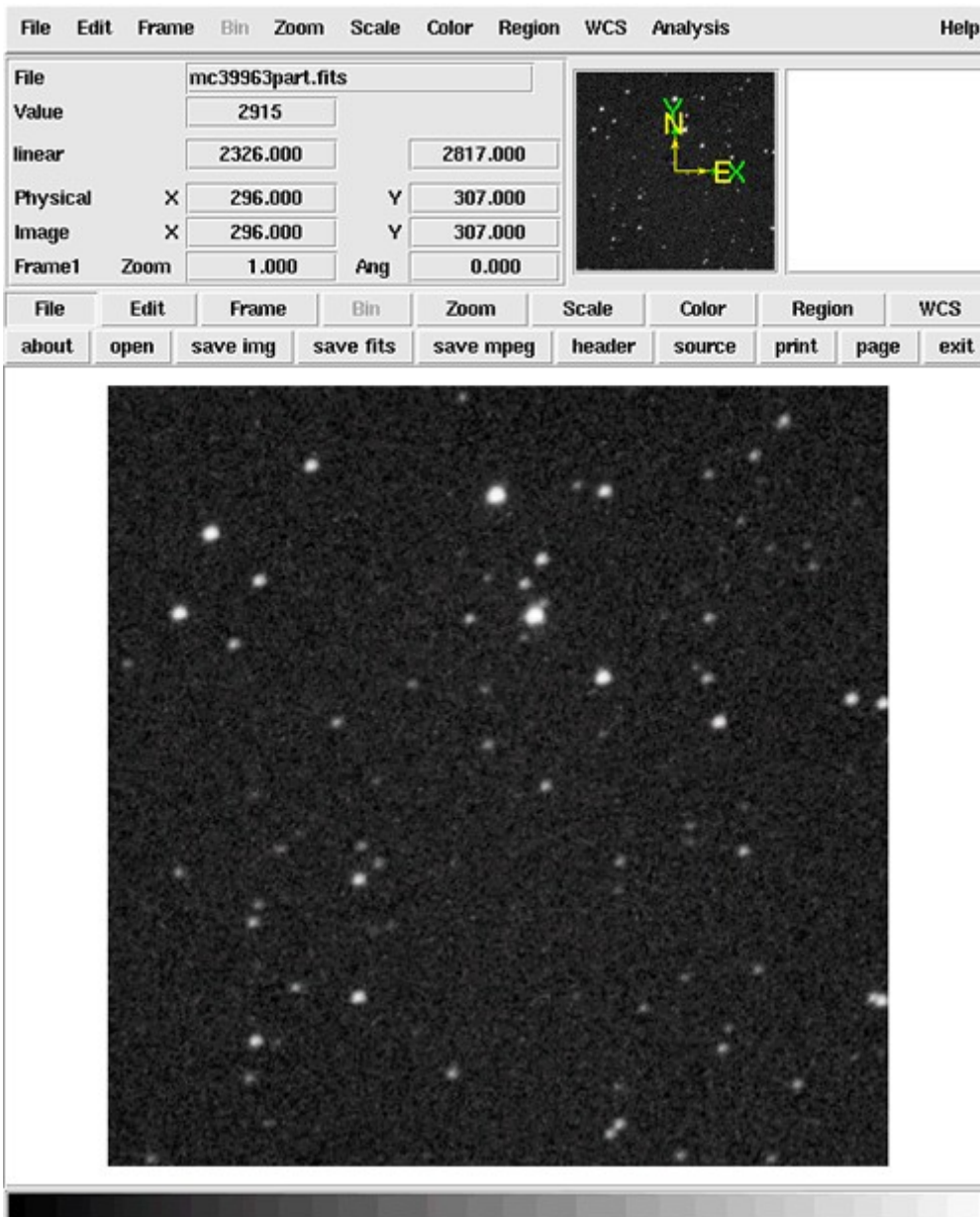
View of entire plate
with each pixel
representing 400
(20x20) pixels in
the scanned image

Access to digital images



The same 1/20 scale image may be downloaded as a FITS image to an image browser such as ds9.

Access to digital images



By clicking on any position in the web display of the entire plate, a portion with dimensions defined on the search page is returned as a FITS or JPEG file at the full scanned image resolution.

Ds9 displays a FITS image of M44 from part of an MC plate.

The Future

- As money becomes available, we will continue to digitize our plates, producing an online image archive and a catalog of objects found in those images.
- We are looking at additional locations to duplicate the petabyte of images we will generate, though we intend to serve it all, too.
- We are imaging our 1200 telescope log books.
- We will continue to transcribe the logbooks listing the positions of the plates we hold.