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

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

- **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
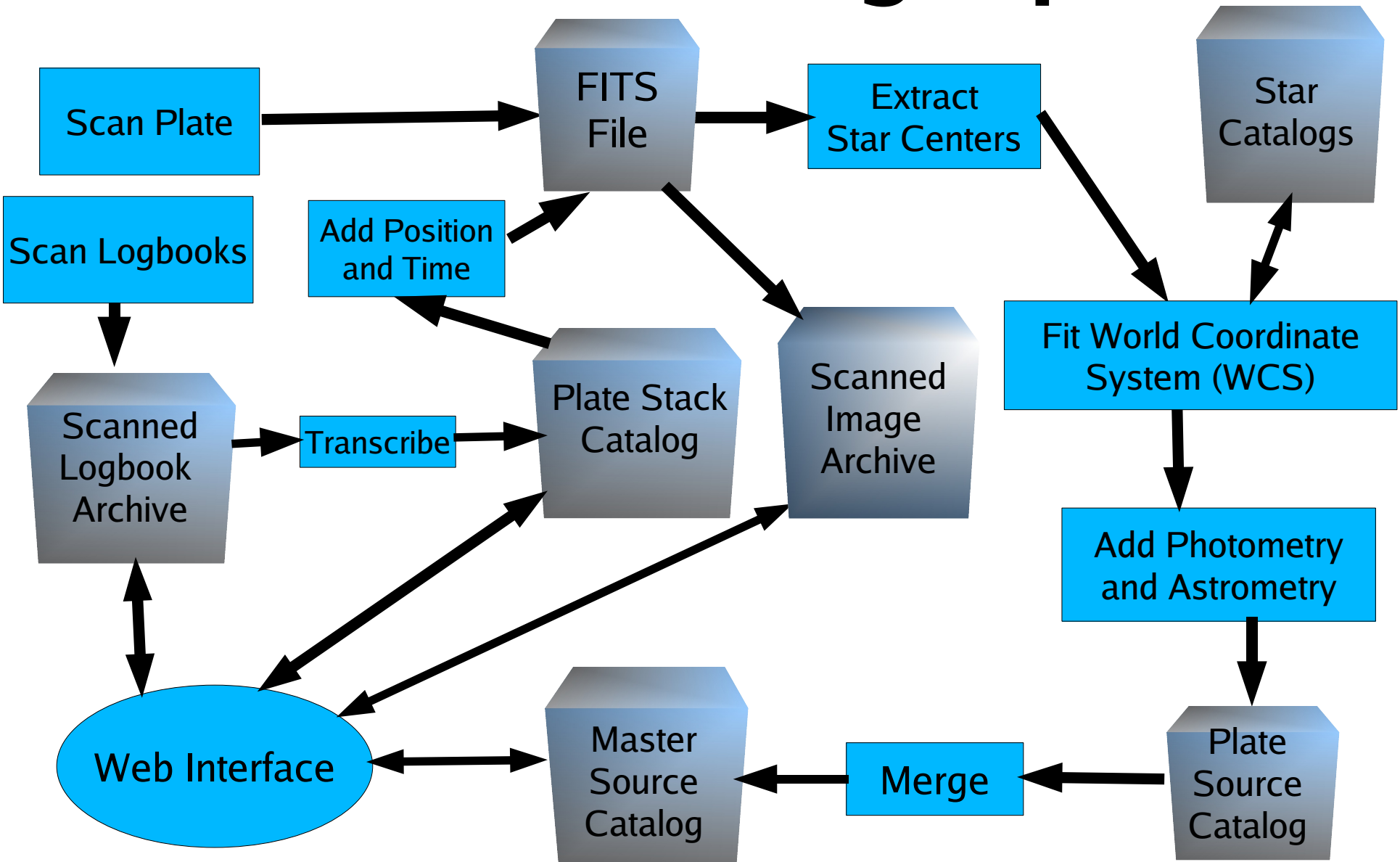
Before We Scan the Plates

- 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
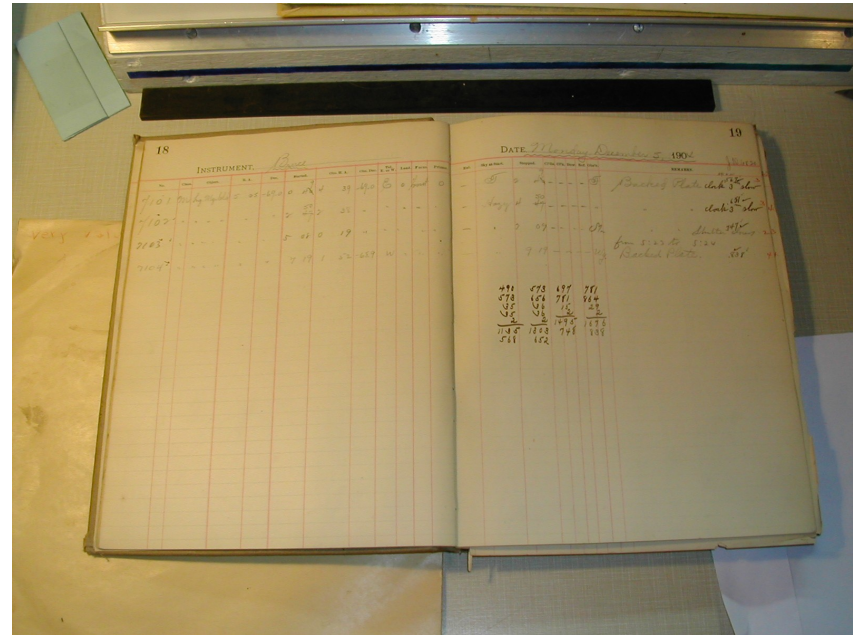
Telescope Log Digitization

- 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-µm 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
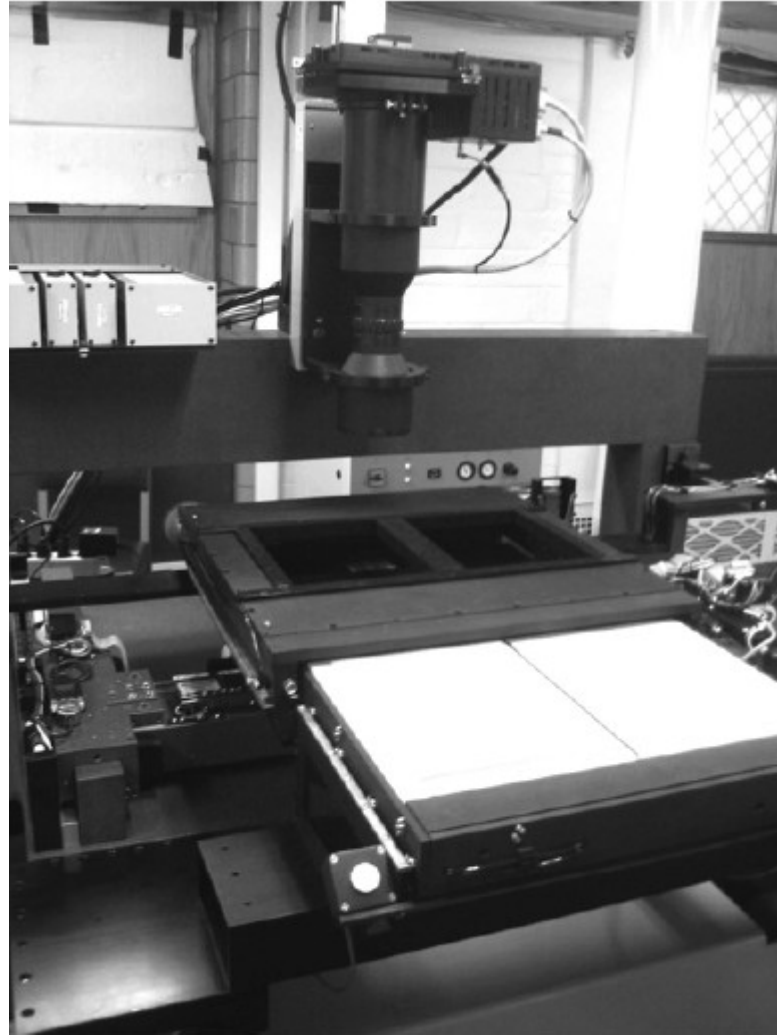
Putting It All Online

• 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
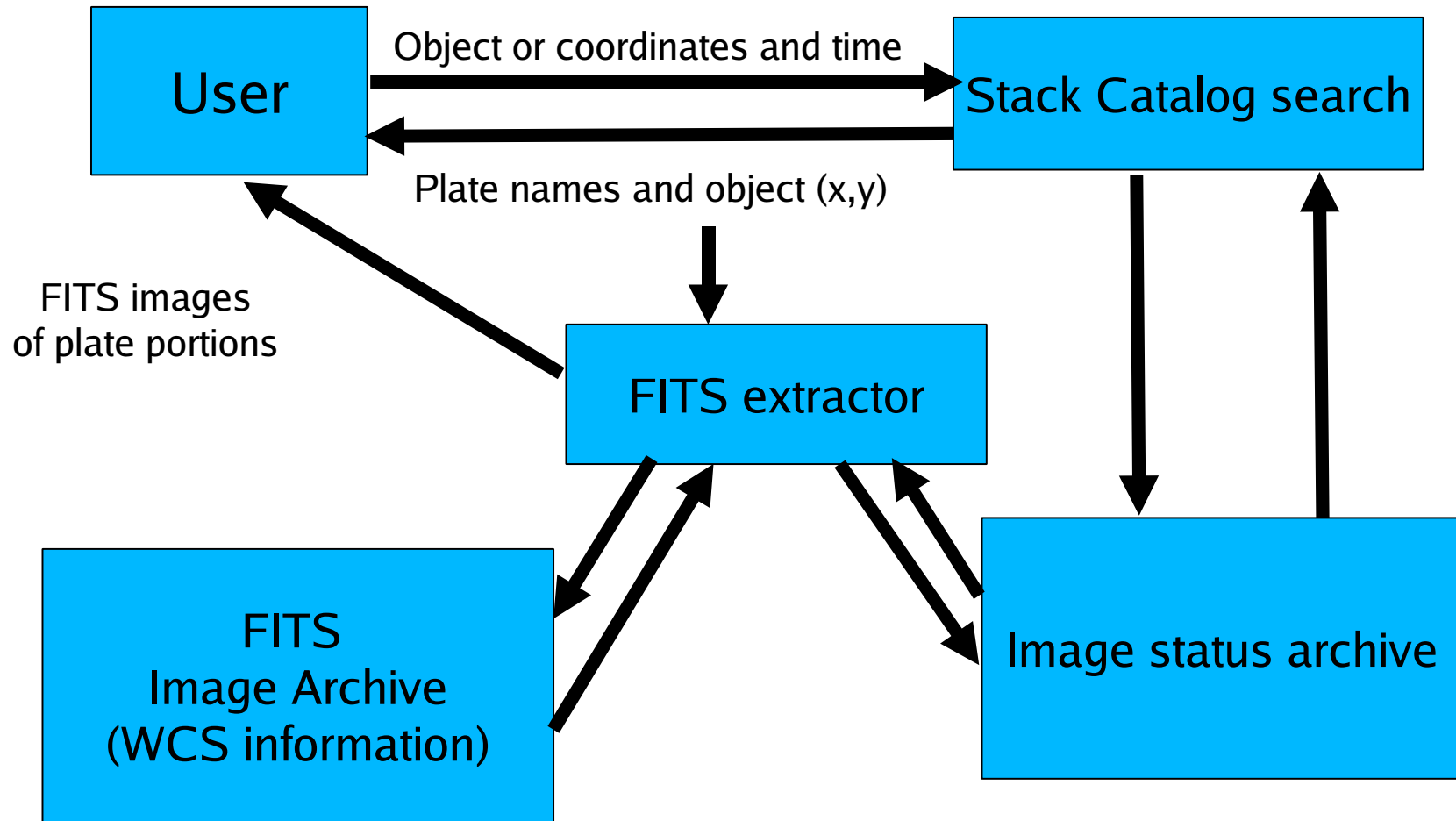
Object or coordinates and time

User → Stack Catalog search

Plate names and object (x,y)

FITS images
of plate portions

FITS extractor

FITS
Image Archive
(WCS information)

Image status archive

Access to digital images

MC Series description page with spatial and temporal distribution

Access to digital images

Search page for MC series

Limits may be set in space and time

Access to digital images

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

Plates which have been scanned may be displayed

Access to digital images
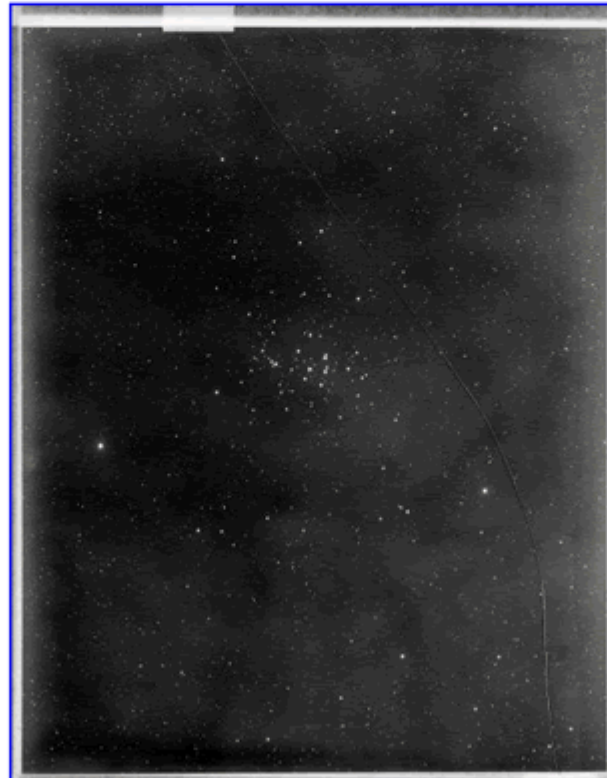

View of entire plate with each pixel representing 400 (20x20) pixels in the scanned image

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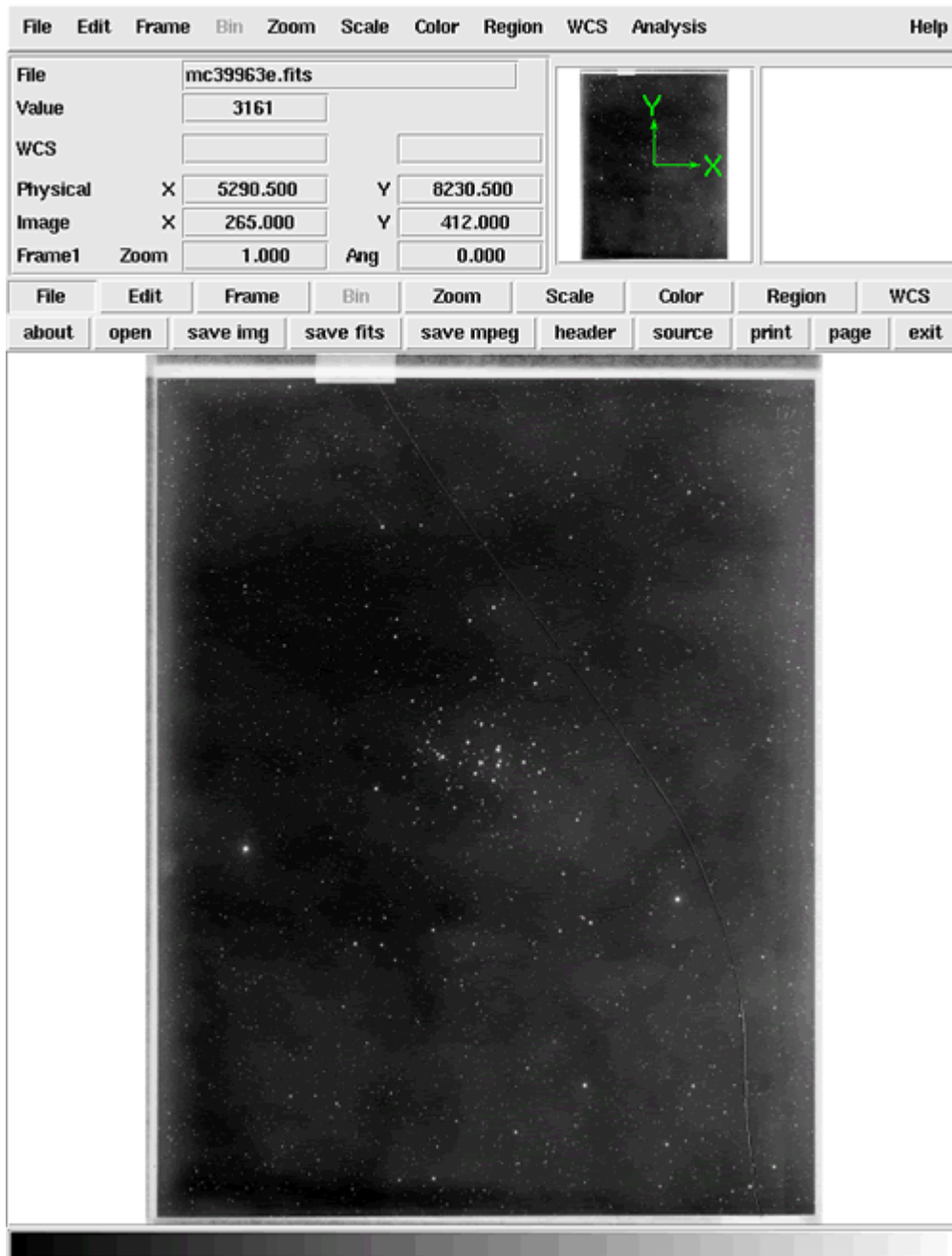
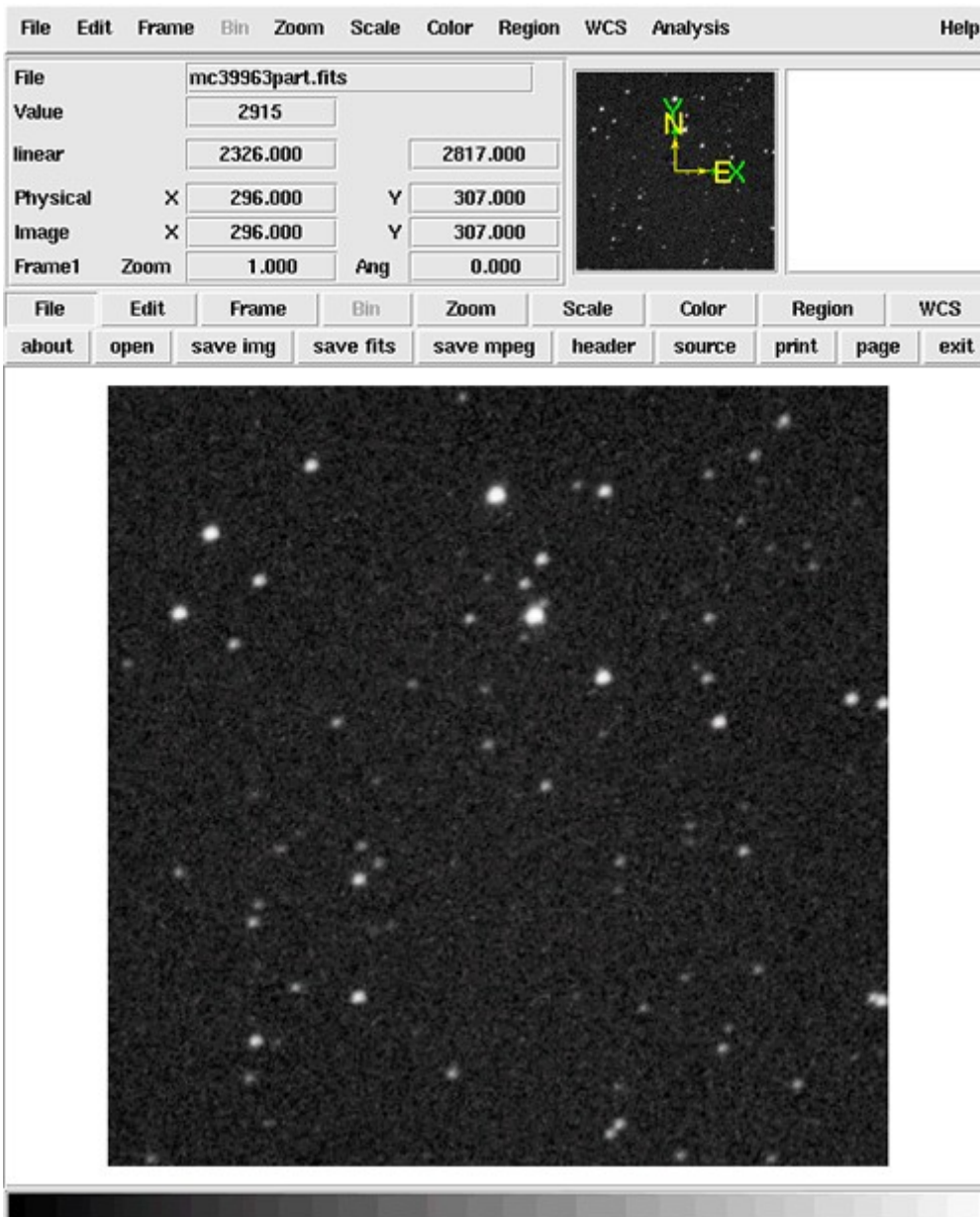The same 1/20 scale image may be downloaded as a FITS image to an image browser such as ds9.

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.