Preserving for Future; Technology Issues

Christina Birdie
Indian Institute of Astrophysics
Bangalore - 34

41st National Conference on “Conservation of Archival, Library and Media Material”, Bangalore 22-24 Jan, 2009

Methods of Preservation

- Preservation
  - Traditional
  - Reprography
  - Digital

IIA Archives Display

- Indian Institute of Astrophysics has more than 5000 items of archival material in the form of hand-written manuscripts, correspondence, printed documents, published Scientific papers, photographs, films, and glass plates, which are subjected to special process of traditional preservation in the last two years.
- While the originals are preserved carefully, IIA is in the process of microfilming the fragile and important documents and papers , especially the annual reports of 18th century which helps us in the process of duplication and digitization for better use without handling the originals.
- Simultaneously we are digitizing all these contents using the Minolta PS 7000, CCD Scanner to convert the analog information into digital in various formats like TIFF, GIFF, JPEG, TXT, RTF,PDF, HTML. They are stored in DVDs and dedicated Server. Those contents which are out of copyright are uploaded in full text form in our open access repository created using open source software Dspace for access in the internet.

Cultural Heritage

Cultural heritage – is defined as objects of historical, scientific, cultural, social, technological, or other value which, regardless of their physical location, are valuable to be preserved for future generations and , using the most expedient technologies must be accessible for public use. Digital cultural heritage is the part of the cultural heritage that consists of objects that are converted to digital format or created in digital format.

(UNESCO Guidelines for the Preservation of Digital Heritage, 2002; http://unesdoc.unesco.org)

Digital Preservation

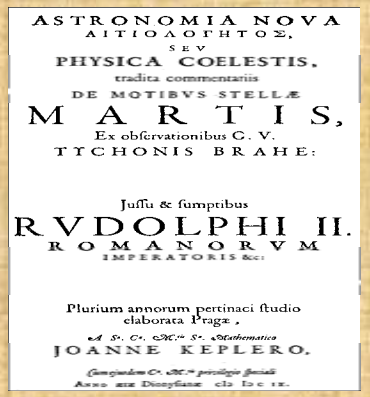
A process which embraces all actions that can be taken with the aim of ensuring the current and long-term survival and accessibility of the physical form, informal content and relevant metadata of archival records, including actions taken to influence records creators prior to acquisition or selection.

(National Archives of Australia (2001))

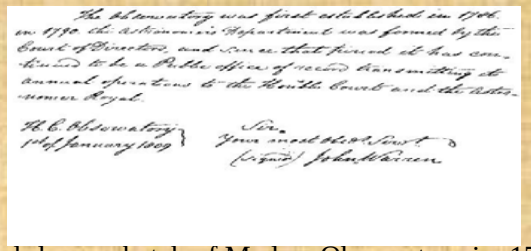
Document Heritage

- Document Heritage
  - Textual documents
  (Manuscripts, posters, books, news papers, journals)
  - Non-textual documents
  (Drawings, maps, prints, music, tapes etc…)
  - Audiovisual documents
  (Photographs, films, discs, tapes etc…)
  -Electronic data of all types in analogue or digital form
  - Web pages, E-Mail messages

Astronomia Nova by Kepler, 1609(Oldest Book in the IIA Library


ASTRONOMIA NOVA
AITIOΛOΓHTOΣ,
SEV
PHYSICA COELESTIS,
tradita commentariis
DE MOTIBVS STELLÆ
MARTIS,
Ex obſervationibus C. V.
TYCHONIS BRAHE:
Juſſu & ſumptibus
RVDOLPHI II.
ROMANORVM
IMPERATORIS &c:
Plurium annorum pertinaci ſtudio
elaborata Pragæ,
A S. C. M.ti S. Mathematico
JOANNE KEPLERO.

Early Manuscript which confirmed the year of establishment of Madras Observatory as 1786



Hand-drawn sketch of Madras Observatory in 1792

Administration of Digital Conversion

- Cost factor for Digital Conversion
  - Cost of digitization depends on the magnitude of the project
    - Hardware
    - Software
    - Networked Access
    - Personnel
    - Dedicated Space
    - Long Term Access
- Technical requirements for Digital Conversion

Media Type	Conversion Method	Resolution	Archive File Format	Screen Presentation Format	Print Presentation Format
Black & White Text Document	Flatbed Scanner or Digital Camera	1-bit, 600 dpi	TIFF w/CCITT Fax 4 Compression	GIF, 4-bit, 120 to 200 dpi	Acrobat (PDF), 1-bit, 300 or 600 dpi
Illustrations, Maps, Manuscripts, etc	Flatbed Scanner or Digital Camera	8-bit grayscale or 24-bit color, 200 to 300 dpi	TIFF	Multiple JPEG, 24-bit, 512x768, 1024x1536, 2048x3072, Quality Level 50	JPEG, 24-bit, 2048x3072, Quality Level 50-100
35mm Black&White & Color slide or negative	PhotoCD or Slide Scanner	24-bit, 2048x3072	PhotoCD or TIFF	Multiple JPEG, 24-bit, 512x768, 1024x1536, 2048x3072, Quality Level 50	JPEG, 24-bit, 2048x3072, Quality Level 50-100
Medium to Large Format photograph, slide, negative, transparency or color microfiche	ProPhotoCD or Drum Scanner	24-bit, 4096x6144	PhotoCD or TIFF	Multiple JPEG, 24-bit, Quality Level 50	JPEG, 24-bit, 4096x6144, Quality Level 50-100
Black & White Microfilm	Microfilm Scanner	1-bit 600 dpi	TIFF w/ Fax 4	GIF, 4-bit, 120 to 200 dpi	PDF, 1-bit, 300 or 600 dpi
		8-bit, 300 dpi	TIFF	GIF, 8-bit 120 to 200 dpi	PDF, 8-bit, 300 or 600 dpi

- Initially digitization in IIA was undertaken as part of the Million Book Project (MBP). The cost involved in digitizing the collection in IIA was absorbed as part of the Library budget, which included the salary for the trainees. The project MBP, donated a Minolta PS 7000 CCD scanner to the library to carry out the digitization programme.

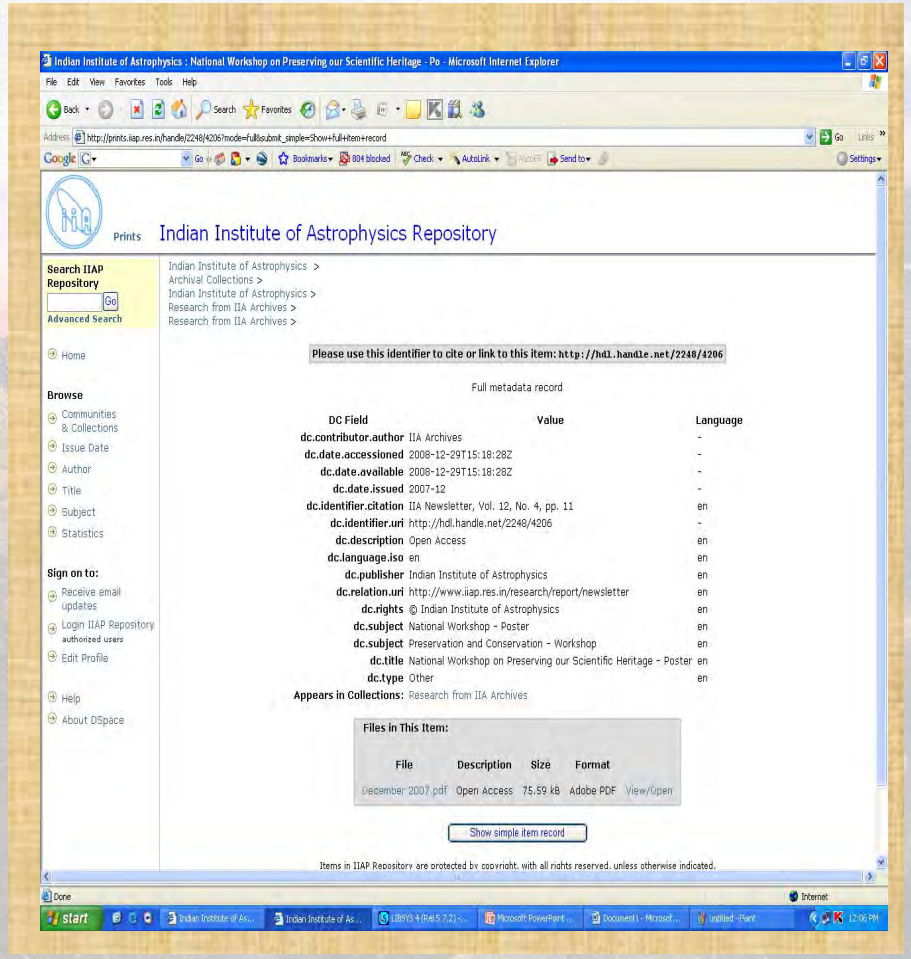
- We have chosen the file formats TIFF, GIFF, JPEG, TXT, RTF, PDF, HTML to save the information and stored them in DVDs. The backup is also stored in the dedicated servers.

Digitization is no doubt the issue that most fascinates and haunts preservation managers in archives and libraries at the moment. The possibilities seem limitless, the advantages are obvious, and from all sides there is pressure to exploit the new medium for preservation purposes, sometimes to the extent that funds are earmarked for digitization that might previously have been allocated to microfilming or conservation…

(http://www.clir.org/pubs/reports/digpres.html)

Metadata

- Metadata
  - The metadata is defined as data about data or information about the digitized object.
  - Categories of Metadata
    - Administrative metadata
    - Technical metadata
    - Metadata on access ability
    - Descriptive metadata



Administration of Digital Preservation

- Cost for Digital Preservation
- Media chosen for preservation
- Standards identified for preservation
  - Open Archival Information System Reference Model (OAIS RM)
  - PREMIS (Preservation Metadata: Implementation Strategies) – a joint collaborative project of OCLC and Research Library Group
  - XML and Schema
  - Persistent Identifier
  - ISO standards for Digital Archiving: ISO 14721:2003 OAIS and RM
- Managing copyright issues
- Consortium of heterogeneous players

- At IIA we have chosen the Open Source Software Dspace to create the OA Repository which includes all the archival digitized contents. http://prints.iiap.res.in/handle/2248/707 .

- We are flexible and open to the process of migration of contents, as and when the technology is upgraded.

- Since IIA holds the copyright for many of the archival items, we have taken the initiative to digitize most of them and also upload them in the repository. Wherever the contents are protected by copyright we have created only metadata for those items.

Digital Preservation Options

- Preservation Methods
- Technology obsolescence protection
  - Refreshing
  - Migration
    - Hardware migration: refreshment
    - Software migration
    - Migration at point of access
  - Normalization
  - Emulation
  - Mirroring
- Analogue Backups

Initially at IIA we had stored the digitized contents in floppy disks which could be read in the earlier generation computers.

As this became obsolete we switch over to magnetic tapes for storing larger volume of data.

As the technology changed we have migrated to CDs and DVDs as the storage medium now.

Physical Storage of Plates, Films & CCDs at Kodaikanal Observatory

Plate Vault Storage

Plates, films & CCDs

Data	Plate		Film		CCD	
	From	To	From	To	From	To
Whitelight	August 1903	September 1975	September 1975	Continuing	March 2008	Continuing
Hα	April 1911	April 1977	April 1977	July 2007		
Ca+K	October 1904	February 1989	February 1989	October 2007	March 2008	Continuing

(Courtesy: S. Muneer, IIA, Bangalore)

DIGITIZERS AT KODAIKANAL OBSERVATORY

- Digitizers are installed at Kodaikanal Observatory to digitize the Photoheliograms, $Ca^{+}K$ and $H\alpha$ spectroheliograms of the Sun obtained from 1904 onwards. The main components of the digitizers are:

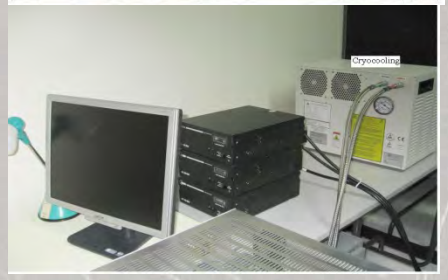
- CCD camera: Format 4K × 4K
- Pixel size 15 micron
- Read out 16 bit
- 4 port read out @500 kHz
- Temperature of operation:
  – 100°C (Cryo cooling)
- Room conditions: Temperature, Humidity and dust controlled

Number of Plates to be Digitized at Kodaikanal Observatory

- White light Images: ~ 44000
- Ca-K line Images: ~ 41000
- H-alpha Images: ~ 38000
- Prominence Images: ~ 34000

(Courtesy: Jagdev Singh et al. Poster displayed during the IIA in-house meeting 15-16 Apr. 2008)

DIGITIZERS AT KODAIKANAL OBSERVATORY



Research projects with Digitized data

- Long term variations in the photospheric and chromospheric rotation rate
- Variation of tilt angle with the solar cycle phase and its implication to the helicity and solar dynamo.
- Variation in the size of the Sun with the phase of solar cycle.
- Variation of background flux with the phase of solar cycle
- Variation in polar regions with the phase of the solar cycle
- Archival of large data base in digital form

(Courtesy: S.Muneer, IIA, Bangalore)

Solar Data Digitization & Administration

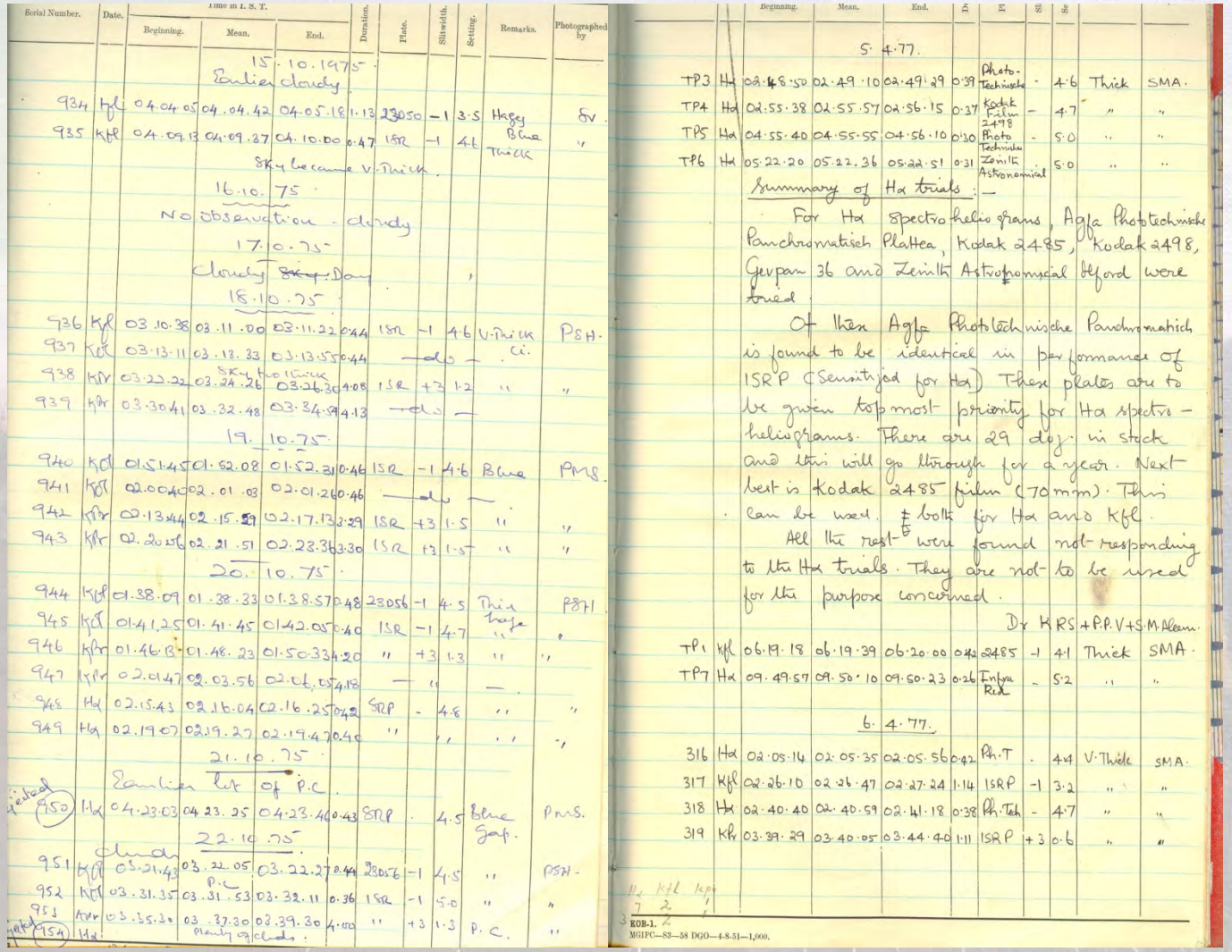
- Data is in ‘astronomy’ FITS format file system
- Individual file name is yymmdd_hhmm.fit where yy is the last two digits of the year of observation, mm is that of month, dd the day, hh is hour of observation and mm is minutes.
- Digitization project will yield 5TB of data
- Data storage is presently done in hard drives and DVDs
- Centrally managed Backups and storage systems are being planned
- Data Access will be enabled from web

File Header

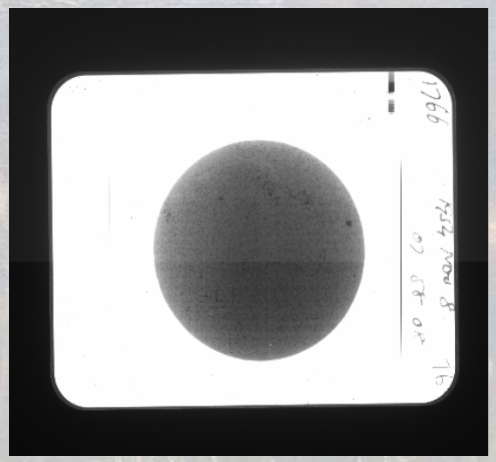
A typical file header looks like this:

980605_0755.fit[4096,4096][short]:
No bad pixels, min=0., max=0. (old)
Line storage mode, physdim [4096,4096], length of user area 284 s.u.
Created Thu 21:49:40 15-Jan-2009, Last modified Thu 21:49:15 15-Jan-2009
Pixel file "980605_0755.fit" [ok]
FITSDATE = '2008-7-30' / DATE FITS FILE WAS CREATED
DATE-OBS = '2008-7-30' / DATE OF EXPOSURE
TIME-OBS = '15:32:45' / TIME OF EXPOSURE
EXPTIME = 5.000000 / EXPOSURE TIME IN SECONDS
INSTRUME = 'SATURN' / DETECTOR USED
SERIALNO = '06370054' / DETECTOR SERIAL NUMBER
PIXELSIZ = 15 / DETECTOR PIXEL SIZE IN MICRONS

Physical Logbook Entry & Digitized Image of the Sun



Digitized Image (White Light)



Digitized Image ($Ca^{+}K$ line)

(Courtesy: S.Muneer, IIA, Bangalore)

Challenges to Digital Preservation

- Storage Media problem
- Hardware Obsolescence
- Software and format Obsolescence
- Web Archiving, unknown files, community data sets

Future Strategies

- Access to data tomorrow requires decisions concerning preservation today
- Adequate funding to support persistent digital access
- Well-defined roles, responsibilities of the stake holders
- Corporate/public partnership with Government initiatives