

**The Preservation of Landsat Data
by the National Land Remote Sensing Archive**

John E. Boyd
U.S. Geological Survey
EROS Data Center
Sioux Falls, SD 57198

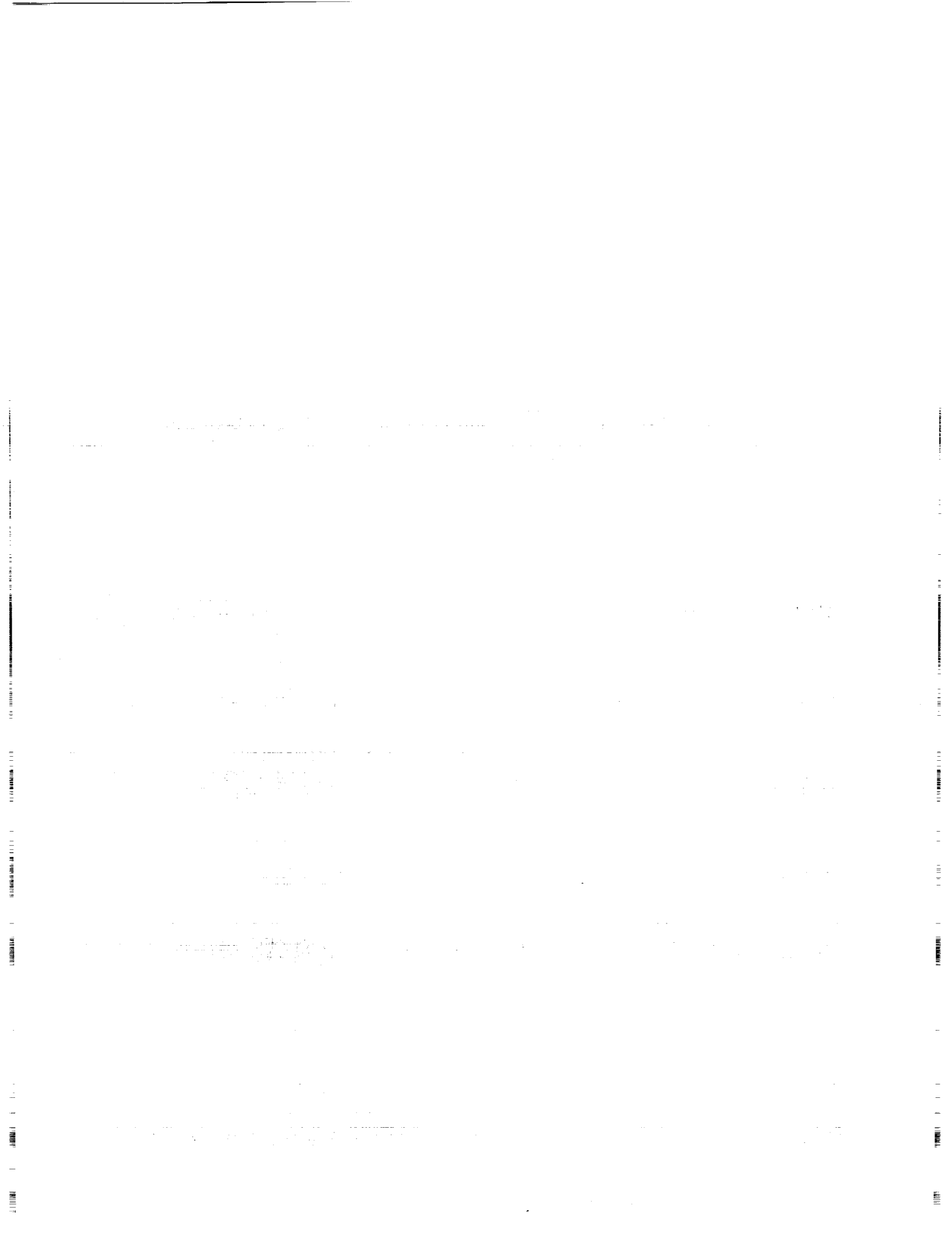
ABSTRACT

Digital data, acquired by the National Landsat Remote Sensing Program, document nearly two decades of global agricultural, environmental, and sociological change. The data have been widely applied and continue to be essential to a variety of geologic, hydrologic, agronomic, and strategic programs and studies by governmental, academic, and commercial researchers. Landsat data have been acquired by five observatories that use primarily two digital sensor systems. The Multispectral Scanner (MSS) has been onboard all five Landsats, which have orbited over 19 years; the higher resolution Thematic Mapper (TM) sensor has acquired data for the last 9 years on Landsats 4 and 5 only. The National Land Remote Sensing Archive preserves the 800,000 scenes, which total more than 60 terabytes of data, on master tapes that are steadily deteriorating. Data are stored at two locations (Sioux Falls, South Dakota and Landover, Maryland), in three archive formats.

The U.S. Geological Survey's EROS Data Center has initiated a project to consolidate and convert, over the next 4 years, two of the archive formats from antiquated instrumentation tape to rotary-recorded cassette magnetic tape. The third archive format, consisting of 300,000 scenes of MSS data acquired from 1972 through 1978, will not be converted because of budgetary constraints.

The data consolidation and conversion project will transcribe approximately 55,000 reels of high-density tape to 1,500 cassettes, ensuring that the data will be readable for the next 10 years. Some of these data, less than 10 percent, will not be reproducible because of the deterioration of the magnetic coating binder, the physical edge damage to the tape, or the demagnetization of the recorded data. The archive conversion activity will involve several computation and data manipulation tasks, in addition to rerecording the data. For example, to ensure that the archive inventory accurately specifies the location and quality of every retrievable scene, data will be spatially referenced after they are transcribed. Transcribed images will be assessed visually for cloud cover and data quality, and a numerical rating will be entered into a land information data base along with other catalog metadata, such as acquisition date and time, solar illumination angle, latitude and longitude, and sensor gain. A subset of each scene will be created during transcription by sampling every 64th element of the original data. The resulting archive of 500,000 browse images (325 gigabytes of data) will be made available to researchers through an interactive land information system.

This data preservation project augments EDC's experience in data archiving and information management, expertise that is critical to EDC's role as a Distributed Active Archive Center for the Earth Observing System, a new and much larger national earth science program.





LANDSAT DATA PRESERVATION

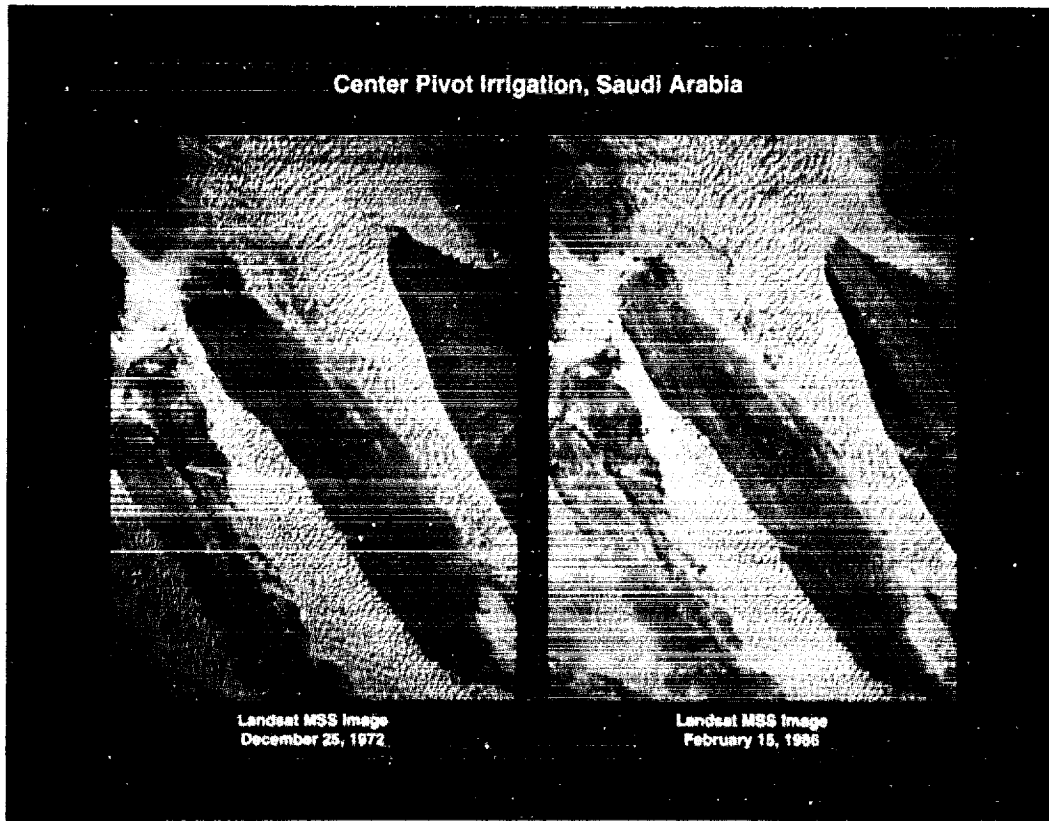
Conversion of Landsat Thematic Mapper and Multispectral Scanner Data in the National Satellite Land Remote Sensing Data Archive

Conference on Mass Storage Systems and Technologies
for Space and Earth Science Applications
July 23-25, 1991

John E. Boyd
EROS Data Center
U.S. Geological Survey

EDSP08 - 12 - 91
TMACS

USGS/EDC
T10





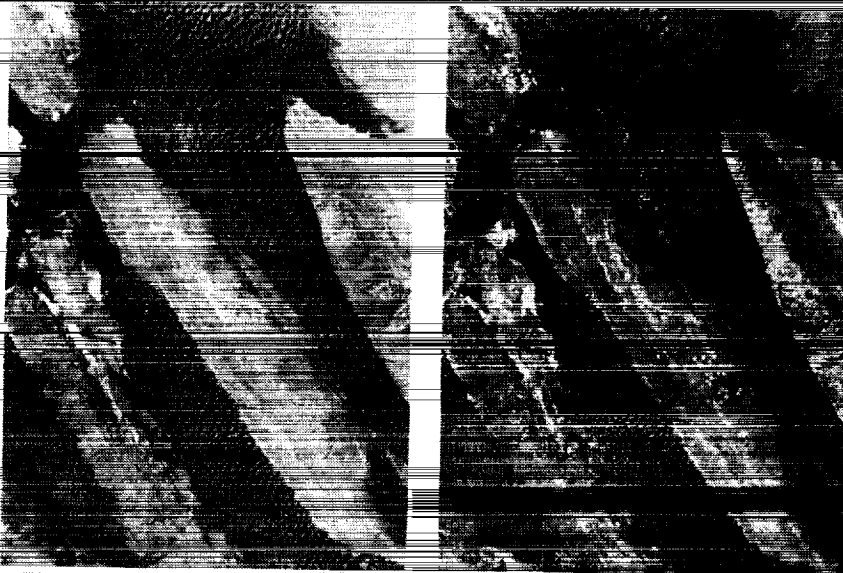
EDC Land Information System Impacts

- Approximately 900,000-scene catalog of U.S.-held worldwide Landsat coverage
- Additional 2,000,000-scene catalog of International ground station TM/MSS data, compiled from contributions from 11 of 16 worldwide Landsat ground stations
- Catalog includes guide, inventory, and fairly extensive metadata
- Online digital browse data will be added to a Global Land Information System (GLIS) currently under development, as archive conversion operations progress
 - ~350 GB of TM digital browse data (3 bands, ~1/64 subsampling)
 - ~460 GB of MSS digital browse data (3 bands, ~1/36 subsampling)
- Currently, digital browse are planned to be archived in two optical disk jukeboxes; they will be available to IBM-compatible PC and Unix workstation inquiry and display

ED8P08 - 12 - 91
TMACS

USGS/EDC
TIO

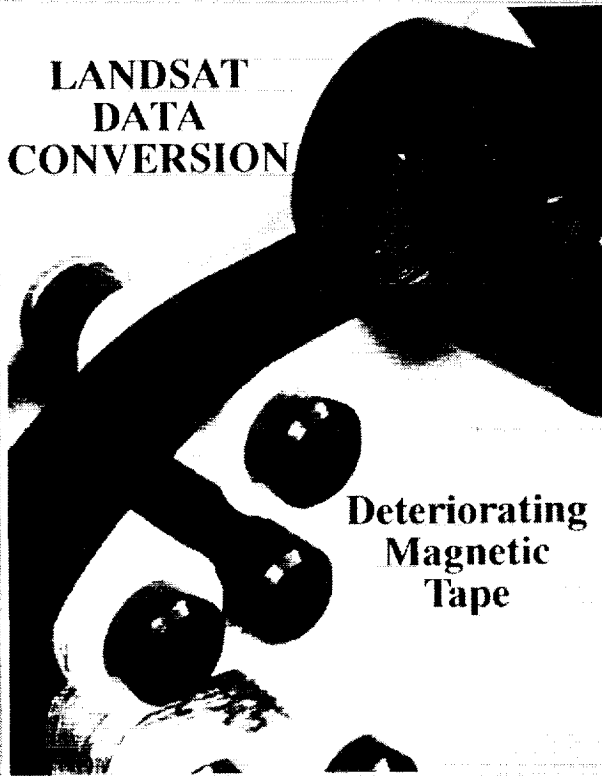
Center Pivot Irrigation, Saudi Arabia



Landsat MSS image
December 25, 1972

Landsat MSS image
February 15, 1986

LANDSAT DATA CONVERSION



**Deteriorating
Magnetic
Tape**

LANDSAT ARCHIVE

MSS data (1972-present)
800,000 scenes
21 terabytes

TM data (1982-present)
170,000 scenes
40 terabytes

REEL TO REEL TAPE



**DURABLE
MEDIA**

MEDIA CONVERSION WILL:

- PRESERVE DATA
- REDUCE PHYSICAL ARCHIVE SPACE BY MORE THAN 90%

NMD 1/9/91



Conversion Objectives

- Transcribe data from high density tapes (HDTs), readable only by scarce obsolete longitudinal-track instrumentation recorders, to cassette-format rotary-head recorders
- Transcribe HDT data to new media before known physical degradation of master tapes results in significant (and intolerable) loss of irreplaceable scene data
- Maintain data integrity during transcription to 1 bit error per 100,000,000 bits recorded, by incorporating deeply interleaved Reed-Solomon error correction code
- Maintain long-term data integrity by using a fully enclosed cassette to protect the 1-inch-wide tape from edge scalloping, which too-frequently occurs on open-reel HDT
- Shrink the physical volume of the archive by a factor of approximately 100, as a result of the high media density and the stacking of multiple HDTs on one cassette
- Record 3 physical HDT formats (14- and 28-track) in a single rotary-head format

LANDSAT DATA CONVERSION

LANDSAT ARCHIVE

MSS data (1972-present)
800,000 scenes
21 terabytes

TM data (1982-present)
170,000 scenes
40 terabytes

REEL-TO REEL TAPE

DURABLE MEDIA

Deteriorating Magnetic Tape

MEDIA CONVERSION WILL:

- PRESERVE DATA
- REDUCE PHYSICAL ARCHIVE SPACE BY MORE THAN 90%

NMD 149 91



Project Goals

- **Ensure availability of the 20-year Landsat data archive to:**
 - traditional Federal & State agency programs, researchers, and commercial users
 - International global change research community
 - Earth Observing Systems science investigators
- **Maintain data integrity at the highest possible level through conservative implementation of recording technologies and meticulous archival procedures**
- **Commit to a program of periodic archive transcription to new media, at 5-7 year intervals, to avoid future technological obsolescence and data degradation**



Archives to be Converted or Transcribed

<u>Data Type:</u>	<u>Acquired:</u>	<u>Volume:</u>	<u>Scenes:</u>	<u>Tapes:</u>
MSS-WBVT	'72 - '78	9.5 TB	310,000	26,000
MSS-P	'79 - '80	2.8 TB	65,000	2,500
MSS-A	'81 - '91	7.2 TB	240,000	5,700
TM-A	'84 - '85	3.5 TB	15,000	1,400
TM-R	'82 - '91	40.0 TB	175,000	16,600
MSS-X-CCT	'72 - '78	1.3 TB	43,500	12,500
TM-P-CCT	'82 - '91	2.6 TB	8,700	35,000

EDSPO8 - 12 - 91
TMACS

USGS/EDC
TIO



Comparison of Storage Technologies

COMPARISON OF HIGH-DENSITY STORAGE TECHNOLOGIES						
Technology	(GB)	(¢ / MB)	Cost Imprv	Volume Impr	Archival	Exchange
CCT	0.15	5.0	Reference	Reference	Excel	Excel
3480	0.20	2.5	2 X	5 X	Excel	Excel
14-Tk HDT	2.3	7.0	0.7 X	8 X	Good	None
Optical Disk	6.0	5.0	1 X	25 X	Excel	Fair
DCRSI	48.0	0.3	17 X	300 X	Excel	Poor
Digital D2	75.0	0.1	50 X	475 X	Good??	Poor
Optical Tape	1,000.	0.7	7 X	2500 X	Good??	Poor

EDSPO8 - 12 - 91
TMACS

USGS/EDC
TIO



Capacity Comparison of Familiar Media

The following number of copies of the M/W Dictionary can be stored on these media:

- 2 on a standard 10.5-inch, 6250-bpi computer-compatible magnetic tape
- 9 on one 5.25-inch CD-ROM optical disk
- 35 on one 8-mm Exabyte-type cartridge magnetic tape
- 100 on one 12-inch double-density WORM optical disk
- 800 on one DCRSI-type or medium-sized DD-2 cassette magnetic tape
- (Merriam-Webster 9th New Collegiate Dictionary: 50-MB text + 10-MB illustrations)

EDSPOS - 12 - 91
TMACS

USGS/EDC
T10



Desirable Characteristics of a Digital Cassette Recorder System

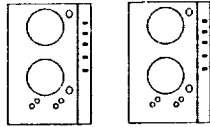
- High data density — 48 gigabytes per cassette
- High data transfer rate — 108 megabits per second
- Acceptable bit-error rate — 1 error in 10^{-8} bits recorded
- Low tape stress & quick rewind — 1,700 ft. rather than 9,200 ft. of tape
- Easy tape handling and no edge scalloping — strong, light polycarbonate cassette
- Archival quality tape formulation — 10-year-life minimum w/ gamma ferric oxide
- Incremental recording and playback — 8-MB I/O buffer permits variable-rate operation
- Rapid data identification and retrieval — IRIG timecode or data-block addressing



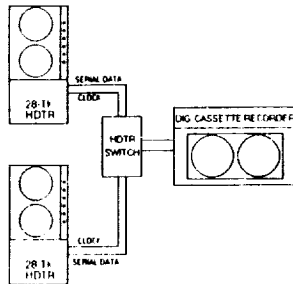
TMACS TM/MSS ARCHIVE CONVERSION SYSTEM

LANHAM HDT DATA TRANSCRIPTION

- 1) HDT ARCHIVE MANAGEMENT
- 2) HDT CLEANING AND PRECISION WINDING
- 3) DATA TRANSCRIPTION TO DIG CASSETTE
- 4) HDT DATA ERROR MONITORING & REWORK



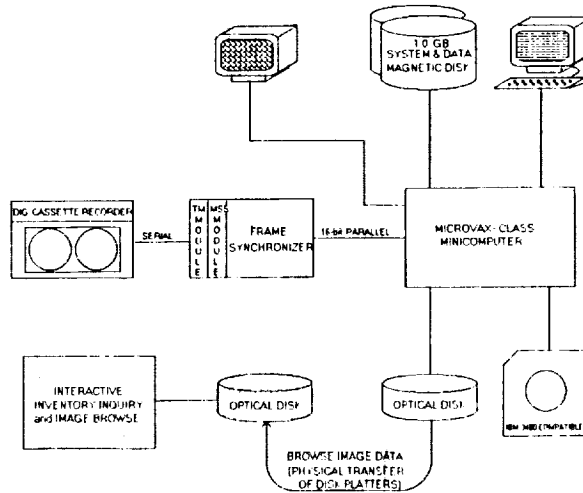
HDT PRECISION CLEANER & WINDERS
(OFFLINE)



EDSP3 - 20 - 91

EDC ARCHIVE / BROWSE GENERATION

- 1) SCENE FRAMING and DATA BASE CREATION
- 2) ANALYTICAL and VISUAL DATA VERIFICATION
- 3) SUBSAMPLED IMAGE BROWSE GENERATION
- 4) ONLINE USER INQUIRY TO OPTICAL DISK LIBRARY



USGS/EDC
T10



Space Compression Resulting from Data Conversion

<u>Sensor</u>	<u>Data Volume</u>	<u>Current No. Tapes</u>	<u>Converted No. Tapes</u>	<u>Current Sq. Ft.</u>	<u>Converted Sq. Ft.</u>
TM	43,500 GB	18,000	1,050	~ 10,000	50
MSS	10,000 GB	8,200	250	1,200	15
MSS-WBVT	9,500 GB	26,000	225	~ 5,000	15
	63,000 GB	52,200	1,525	~ 16,200	80

- **Note: Conversion of MSS Wideband Video Tape (WBVT) currently is neither funded nor scheduled, but USGS vigorously has sought a supplemental FY92 appropriation.**



Archive Conversion — Rates and Volumes

- Assume:
 - TM archive growth of ~25 scenes per day during the next 12 months
 - Final TM archive size to be ~200,000 scenes contained on ~19,000 HDTs
 - 250 workdays per year X 2 years required to complete the archive conversion
 - HDT playback to be ½ realtime rate — 42.5 Mbit/sec, or 5.3 MB/sec
- Then:
 - A production goal of 40 HDTs to be processed per 2-shift day
 - ~420 scenes and 97,000 MB to be transcribed to DCRSI media, daily
 - ~2.2 DCRSI cassettes to be recorded, daily
 - HDT transcription requires: 18,300 seconds (5.0 hours) of capstan time, daily
 - HDT precision clean and rewind requires: 10 hours daily, at 15 minutes per tape

EDSPO8 - 12 - 91
TMACS

USGS/EDC
TIO



Data Conversion Milestone Chart

TASK / ACTIVITY	FY91				FY92				FY93				FY94			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
ARCHIVE CONVERSION SYSTEM PROCUREMENT			CA													
SYSTEM DEVELOPMENT / TEST / INSTALLATION							I/T									
TM ARCHIVE CONVERSION OPERATIONS - LANHAM, MD							R	---	---	---	---	---	R	A		
TM DATA QA & BROWSE GENERATION AT EDC							R	---	---	---	---	---	R	A		
MSS ARCHIVE CONV., DATA QA, & BROWSE GEN. - EDC									A	---	A	P				
1972 - 1978 MSS WBVT ARCHIVE CONVERSION ???																

EDSPO8 - 12 - 91
TMACS

USGS/EDC
TIO