ChannelScience

Detecting the Future of Data Storage SM

The Future Looks at Your Past:

Digital Archeology and the Hard Disk Drive

02006

Diskcon ~~2006~~

Celebrating the 50th Anniversary of the Hard Disk Drive
September 14, 2006

Chuck Sobey
csobey@channelscience.com
972-814-3441
www.ChannelScience.com

What is Archeology?

- Anthropology: The Study of Humanity
  - Our physical characteristics as animals, and our unique non-biological characteristics we call culture.
    - Biological (physical) anthropology
    - Cultural (social) anthropology
    - Archaeology
- Archeology: The Study of Extinct Human Societies using the Material Remains of their Behavior

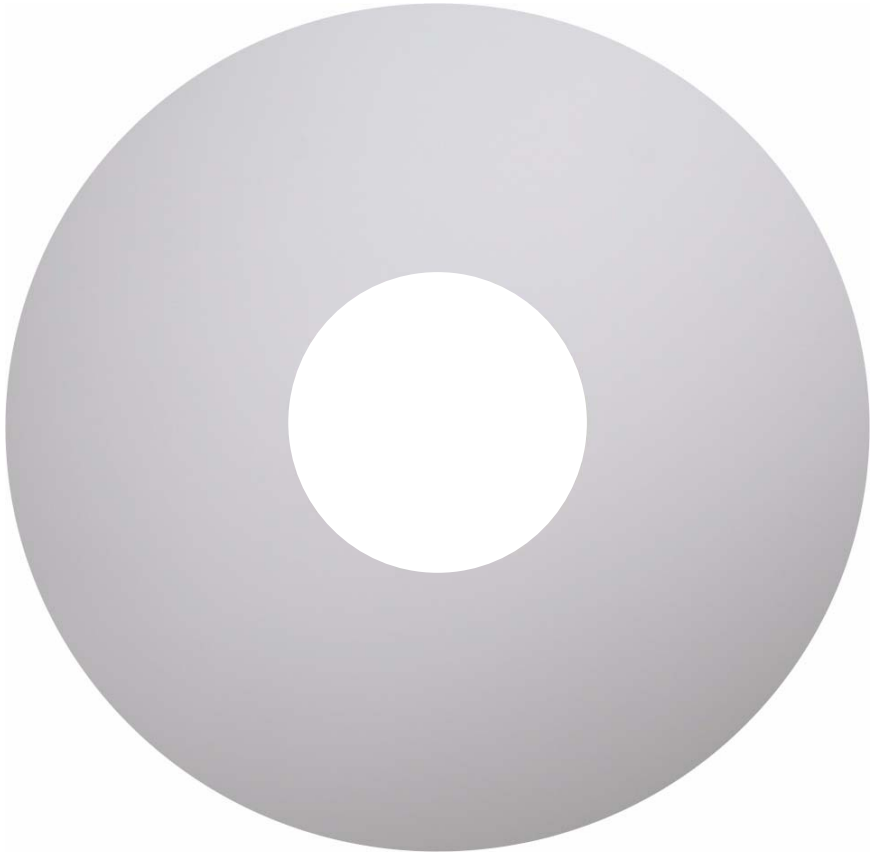
The Material Remains of Our Behavior

Most from www.HitachiGST.com/techtimeline

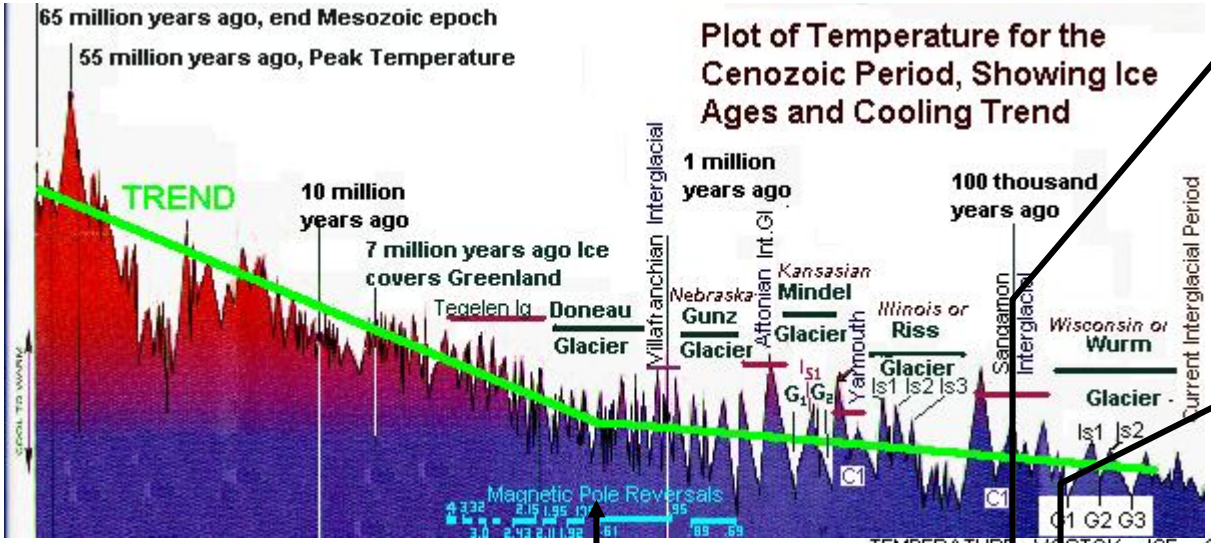
Archeologists: Not Necessarily *Indiana Jones*

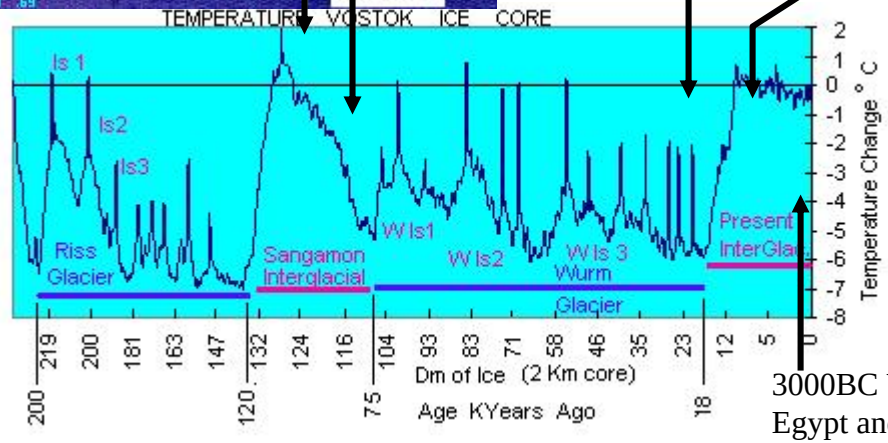
- Philosophers
- Grave Robbers
- Antiquarians
- Scientists (and Engineers)
  - Pseudo-archeologists (Remember *Chariots of the Gods*?)

Loiyangalani, Tanzania ~70,000 years ago



Archeological Timeline and Interglacial Periods





100kybp 1st Modern humans in Africa

Neanderthals in Europe and
Modern humans in SW Asia

Forager societies and
European art tradition

Origins of food production

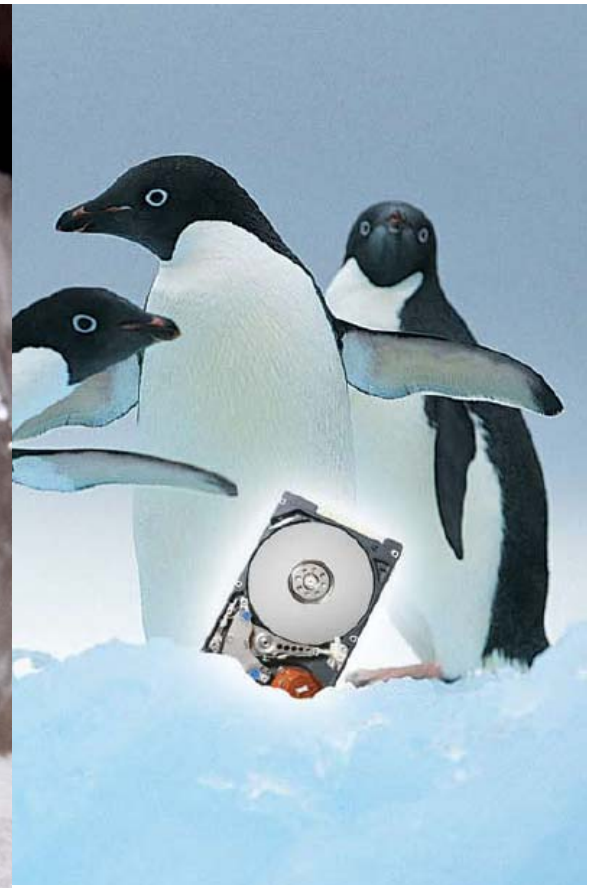
2.5Mybp 1st Toolmaking
Homo habilis and other hominids

1Mybp Fire domesticated
Homo erectus radiates out of Africa

3000BC Writing
Egypt and Sumeria

Images from www.nexialinstitute.com/glaciation.htm
Annotations from B.M.Fagan, *In the Beginning*

The Next Ice Age is 10,000 to 20,000 Years Out



Today's Tools for Future Digital Archeologists

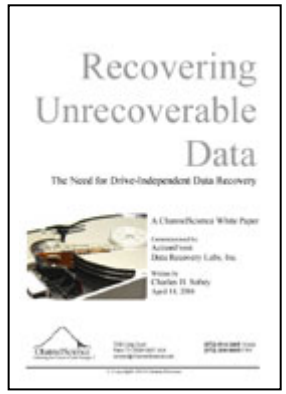
- Data Forensics
- Data Recovery
  - Getting usable data from a failed storage device

Maybe not that failed!

Courtesy Barry Bailey Photography

For More Information on Current Data Recovery Techniques

- Download *Recovering Unrecoverable Data: The Need for Drive-Independent Data Recovery*, Charles H. Sobey



www.ActionFront.com (Seagate)

- C.H. Sobey, L. Orto, and G. Sakaguchi, "Drive-Independent Data-Recovery: The Current State-of-the-Art," *IEEE Transactions on Magnetics*, February, 2006, pp. 188 - 193

The Data Recovery Specialist's Problem

- Deliver contiguous groups of bytes correctly, over the interface, as sectors or logical blocks (LBAs)
- Recover as much file system (directory) information as possible
- Rebuild files (from LBAs) based on directory information or specialized software utilities

The Archivist's (Preservationist's) Problem

- Expect correct bytes are available over the interface
  - File system is also in tact
- Is the application that works with the data still available?
  - Drawing, music, text, computer instructions, …
- Is the operating system that runs the application and controls the data still available?
- Is the hardware that runs the application and operating system still available?
  - Microprocessor
  - Printer, sound card, graphics, …
- Many archivists want to preserve not just the data, but also the look-and-feel of the software application that used the data

The Archeologist's Problem

- The data recovery specialist's problem
- The archivist's problem
- Plus
  - What is a drive?
  - Archeological fallback position
    - Probably had a "religious significance"
  - Why are the drives located where they are?
  - Who had them? Who did not?
  - There's data on it?!
    - Why was the data important enough to be stored?
  - Get funding!

Unattributed Lament

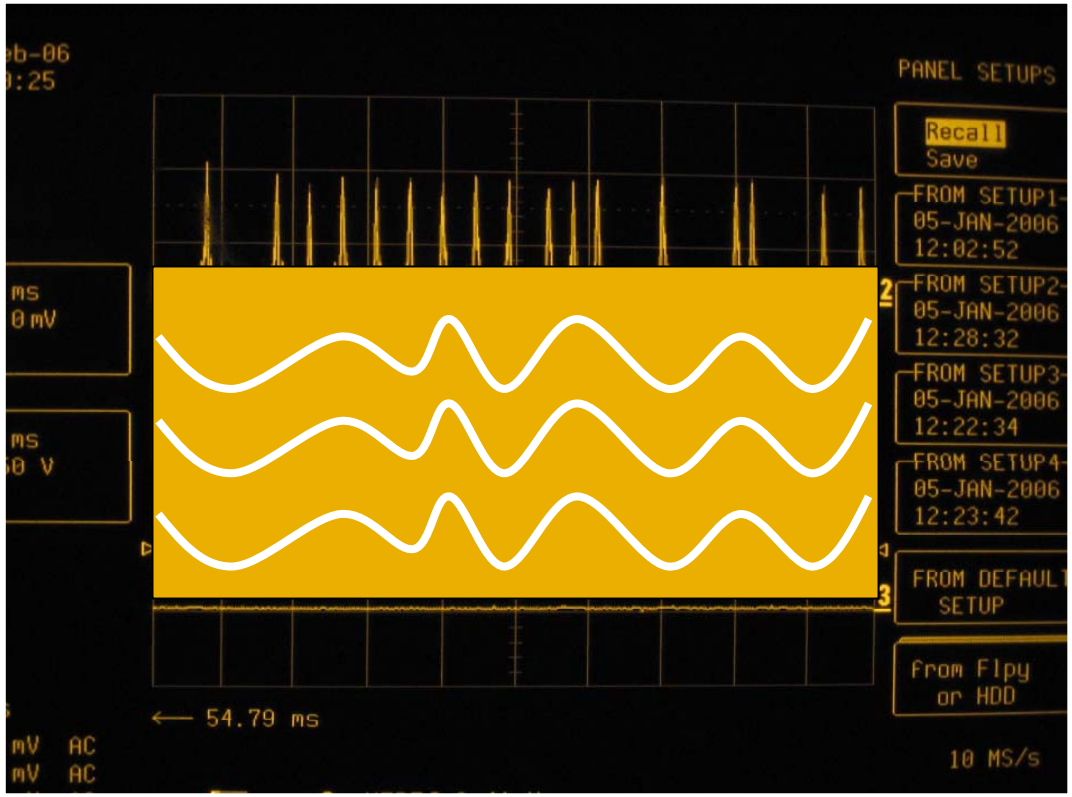
"We are at the beginning of the Digital Dark Ages. A generation's worth of digital data is about to be lost."

"However, there will be plenty of paper evidence left of who's to blame."

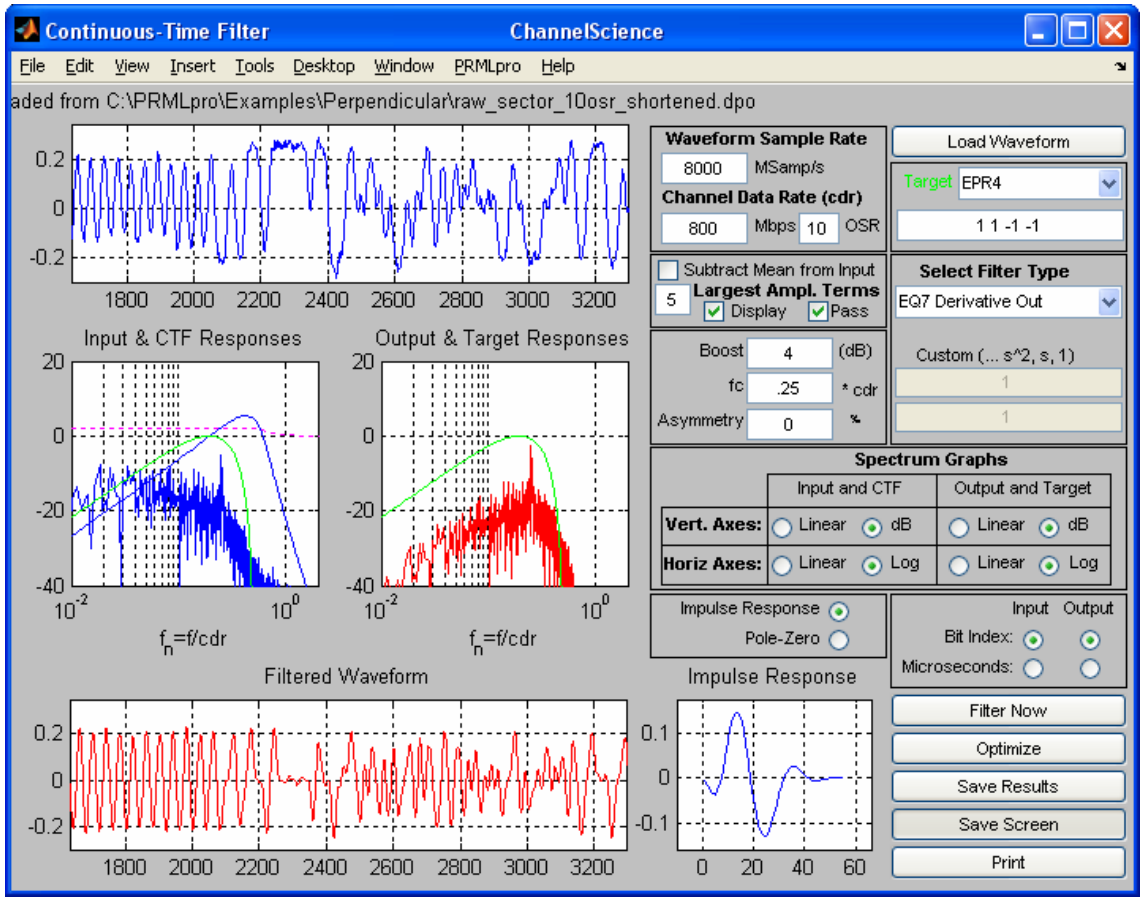
It Won't be Al Hoagland Who is Blamed!

- Readback signal from refurbished RAMAC
- Assuming NRZI
  - 1s are peaks
  - 0s are no-peaks
- Waveform captured February 28, 2006
- Source Al Hoagland

Happy Birthday RAMAC!



Modern Perpendicular Drives Look a Little Different (but NOT on the outside or the inside!)

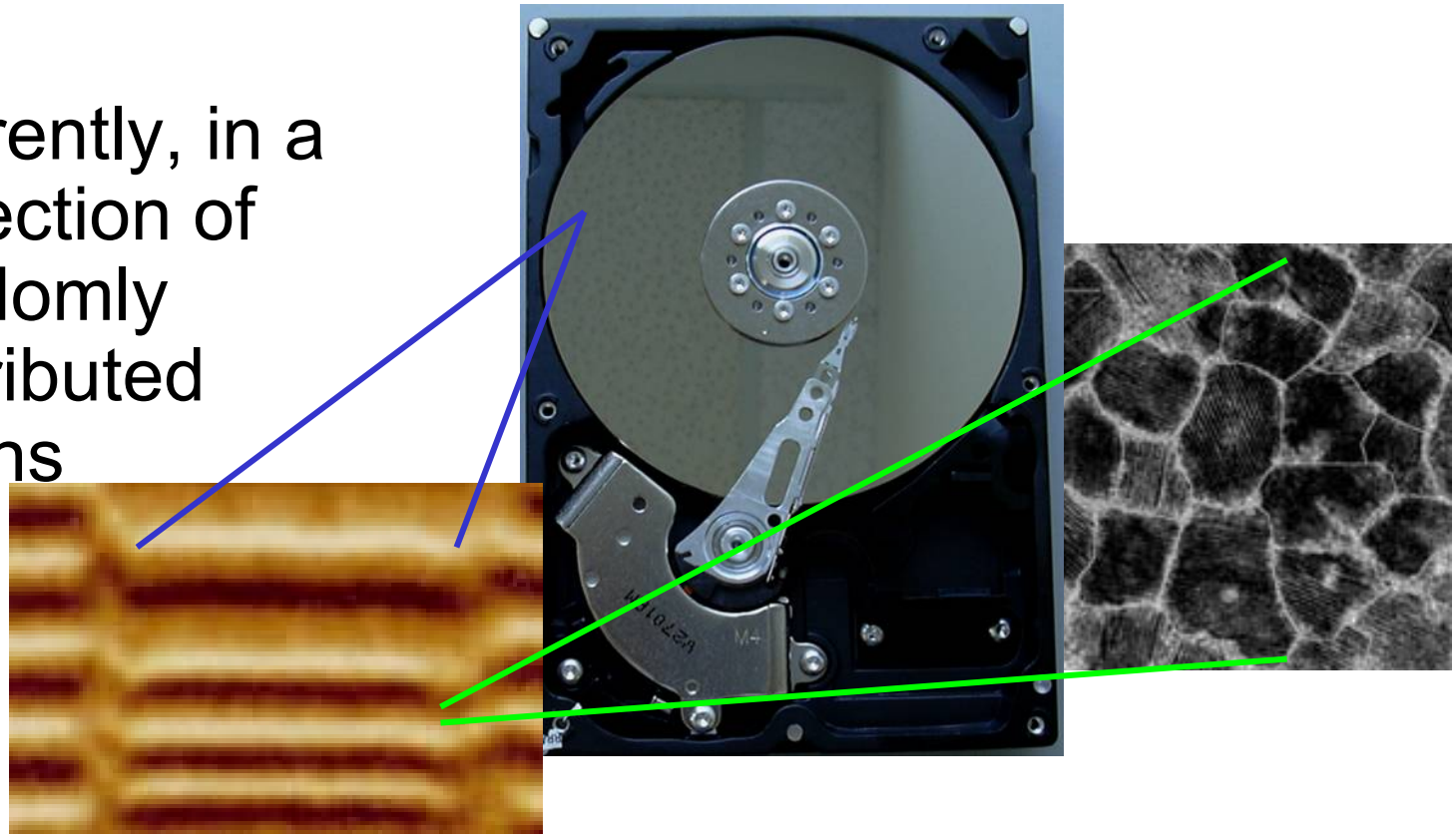


PRML*pro* ™

www.ChannelScience.com

Where do Bits Live?

- Currently, in a collection of randomly distributed grains



Bits are Bits – Not Even Close!

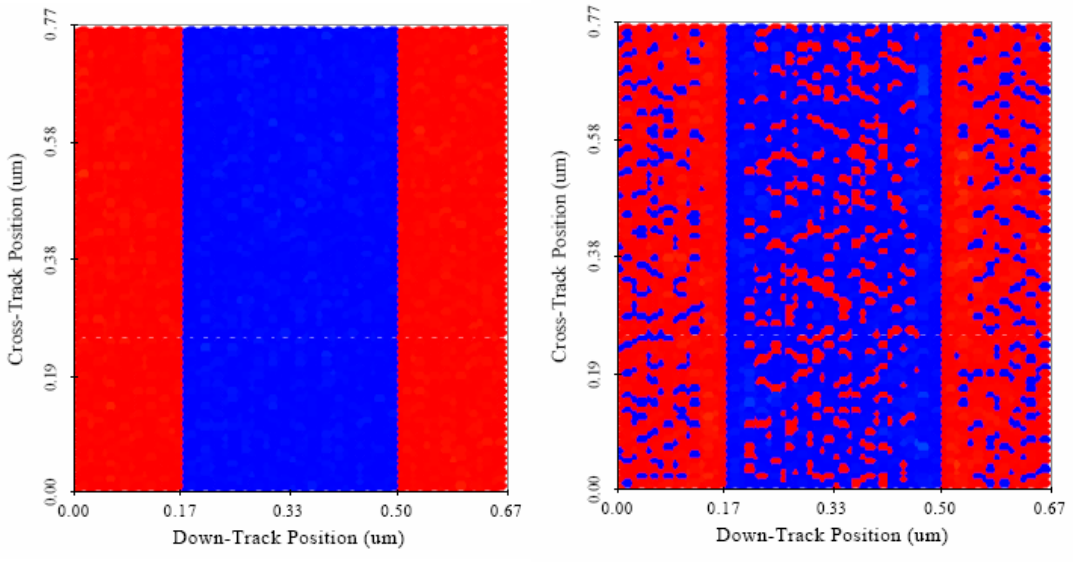
- User bits vs. Channel bits
  - The user's data NEVER get to the disk surface
  - Encoded "channel bits" are stored instead
- Data bits vs. "Maintenance bits"
  - 5 to 15% of the bits are there to help the drive find the data bits
  - If maintenance bits fail, the data bits are likely inaccessible
    - This is a common "repair" in the best data recovery labs today

Good Bits Gone Bad

- Typical unrecoverable error rate spec: 10^{-13} – 10^{-15}
- Degrading signal quality with time, temperature, and stray magnetic fields
- Disk defects, corrosion, thermal asperities, …

Where do Bits Go?

- Perpendicular bits also decay (like longitudinal)
- Perpendicular media support higher densities before the onset of superparamagnetism
- The areas between transitions are the least stable parts of the data



Courtesy Jay Hoinville, Euxine Technologies

How Long can Magnetic Bits Last?

- A rule-of-thumb for the industry
  - No more than 1% readback amplitude decay per “decade” on average
  - Of course a decade ≠ 10 years
    - 1 second; 10 seconds; 100 seconds; 1000 seconds; …
  - 31,536,000 seconds = 1 year
  - 100,000,000 seconds = 8 “decades”
  - 1000 years = ~ 10.5 “decades”
  - 5,000 years = ~ 11.2 “decades”
  - 50,000 years = ~ 12.2 “decades”
- Currently, perpendicularly recorded bits are reported to decay at about 0.1% per decade
- A lot of bits should still be present on a benignly neglected drive

This Isn't a Guarantee – But it is a Possibility!

- Here's the Guarantee
  - If magnetic patterns remain, they WILL contain lots of errors (on the most important bits)
- The Hope
  - Future humans and their technology will be more than capable of deciphering the bits that are there and reconstructing the missing ones – Math is math!
  - They will care enough about us to do it
- Do we care enough about them to make it easier to do?

Gordon Bell: www.MyLifeBits.com

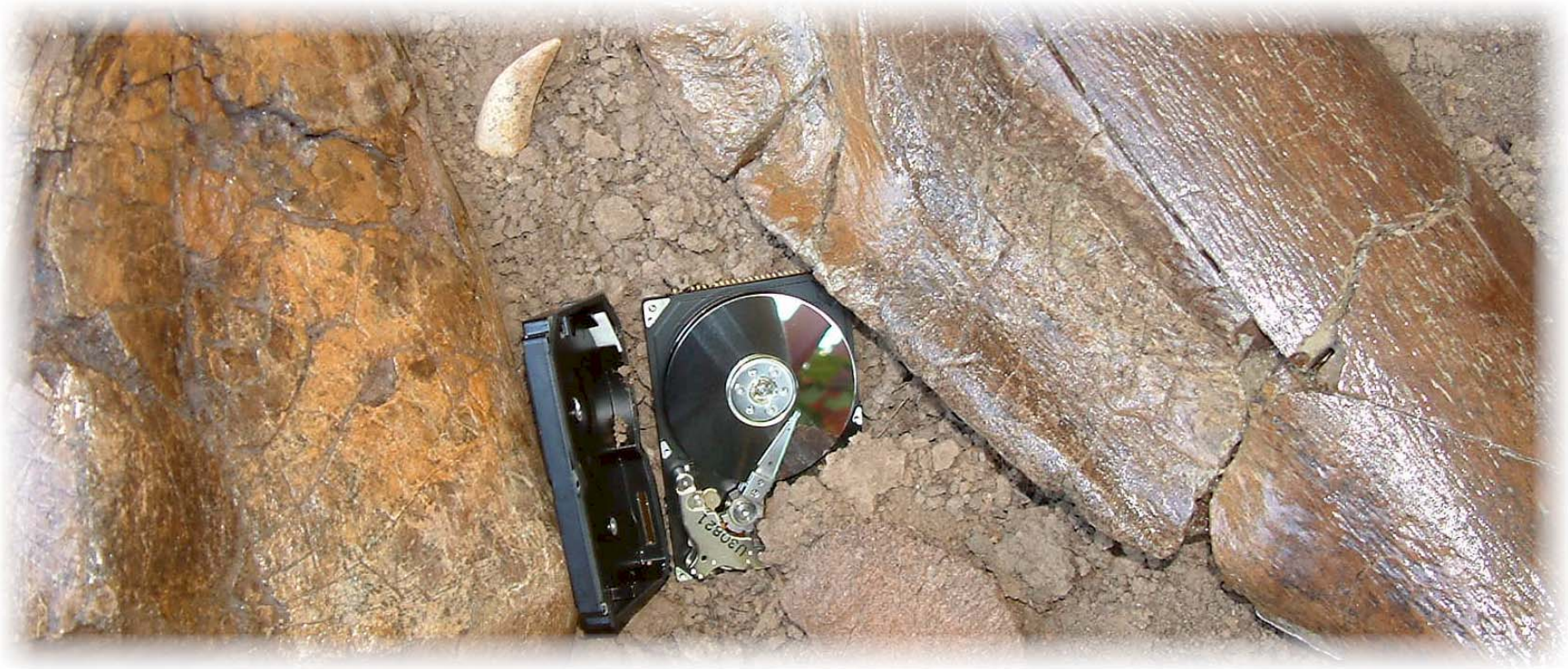
- Microsoft Researcher and key VAX developer
- Recording everything since 1998
- SenseCam camera
- New search capabilities being created

Photo by Mark Richards

The Short-term Future
Grandma and Grandpa's Hard Disk Drives

- We look at old photos and documents now
- We need new search capabilities to find the relevant history and make it accessible (MyLifeBits – Microsoft Research)
- Automatic authors of family histories and personal biographies
  - Random conference paper generator
    - http://pdos.csail.mit.edu/scigen/
  - Automatic generation of ads based on search terms
  - Augment with other contemporary accounts and news reports
  - Improvement: as told by Mark Twain, James Joyce, Hemingway, …
- Synthetic Interview technology from Carnegie Mellon University
  - Albert Einstein and Benjamin Franklin
  - Can "it" continue to learn?
  - Continue to offer advice to future generations
- Ian Pearson – Head of Futurology Unit of British Telecom
  - Could you tell its not the real person over email?
  - Achieve digital immortality
  - By 2050 for the wealthy; 2080 for the rest of us

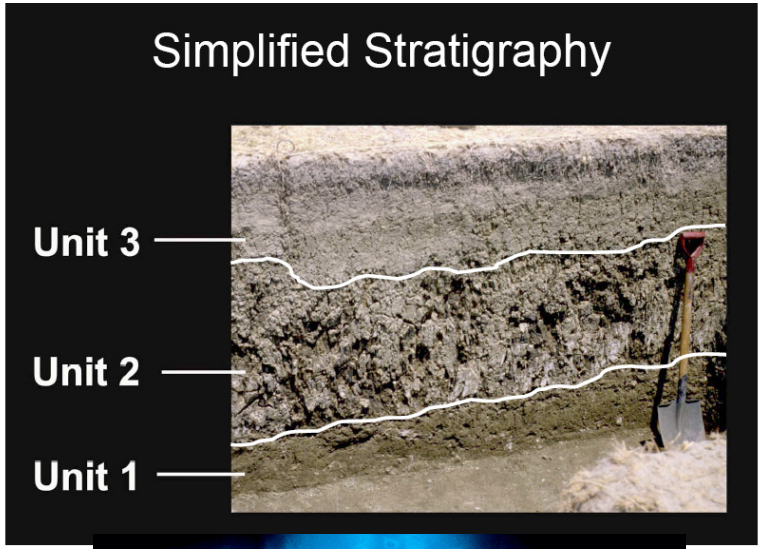
The Long-term Future
An Archeologist Finds a Hard Disk Drive

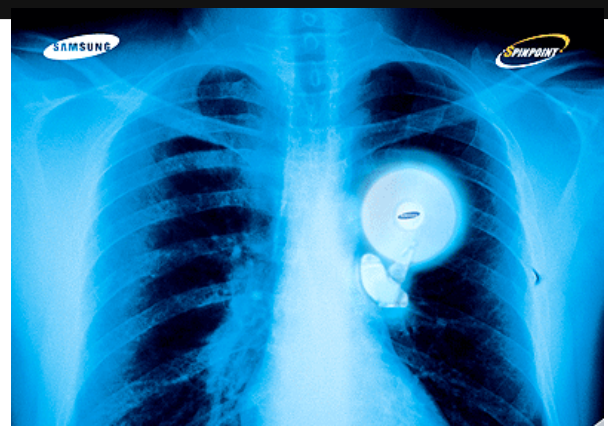


Fossils and photography from the Steven Brittenham collection

An Archeologist Finds a Drive: Now What?

- Record the context of the find
- Are there “written” records identifying the artifact and its purpose?
- External physical exam
- Microscopic analysis (external)
- Internal imaging techniques
- Careful disassembly
  - Trying to get it to work first would probably be a disaster for the data
  - Will they even know it contains data?





What Data May be There?

- Everything – we delete nothing
  - Decision time costs more than storage
- Trans-generational data
  - Passing down your archive to your children
  - Digital essence? (extra-genetic information)
  - Data mutation over time (incorrect copying)
- The steps leading to
  - The “Digital Dark Ages”
  - New methods of doing science, art, government, …
  - The “exaptation” of new capabilities in humans
  - Why there are still no flying cars!

Another Guarantee

- Whatever we have carefully preserved, migrated, and verified for posterity will NOT be what interests the future most

What if No Data Remain?

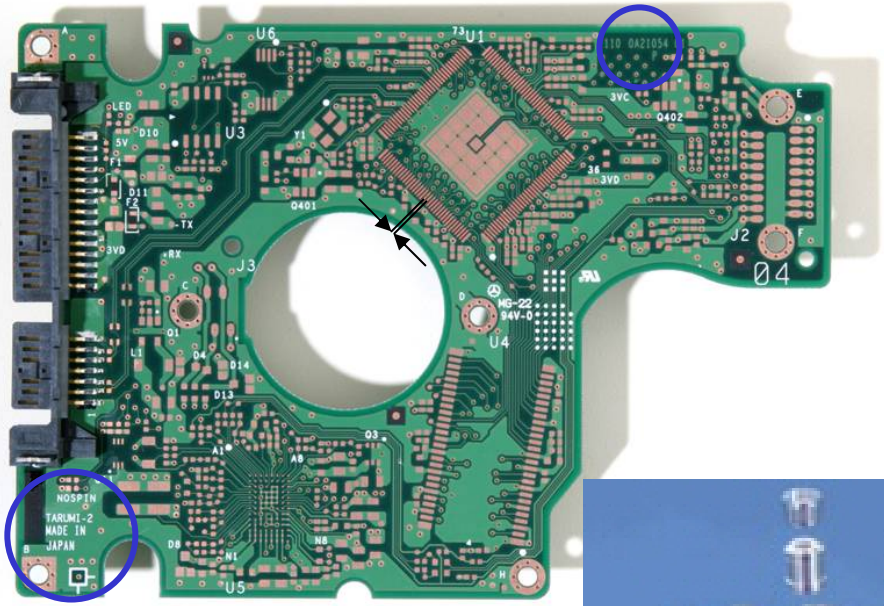
- Drives are very sturdy (paradoxically)
  - Compared with other parts of the PC
  - Beaker People → Hard Disk Drive People
- Who has drives and who doesn't?
- Where are the drives?
  - Geographically
  - In the home, office, school, etc.
- Trace trade routes
  - Development centers; sources of raw materials
  - Which components were made where, and when?
  - See book *From Silicon Valley to Singapore* by David G. McKendrick *et al*
- Manufacturing technology
  - HDD represents a large cross-section of technologies
- Provide chronology

Grave Goods
(of the “Hard Disk Drive People”)

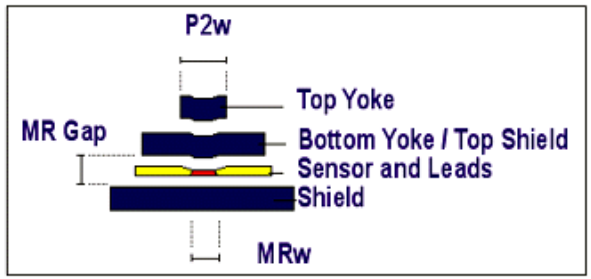
- Spear head
- Hard disk drive

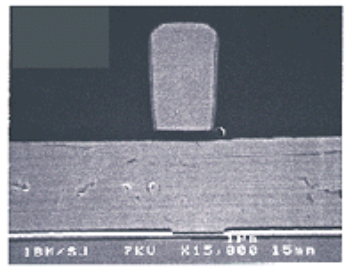
Note: This is a mannequin, not actual human remains – No disrespect!

Manufacturing
Location, Date Codes, Technology



Critical Features of Magnetoresistive Thin Film Head





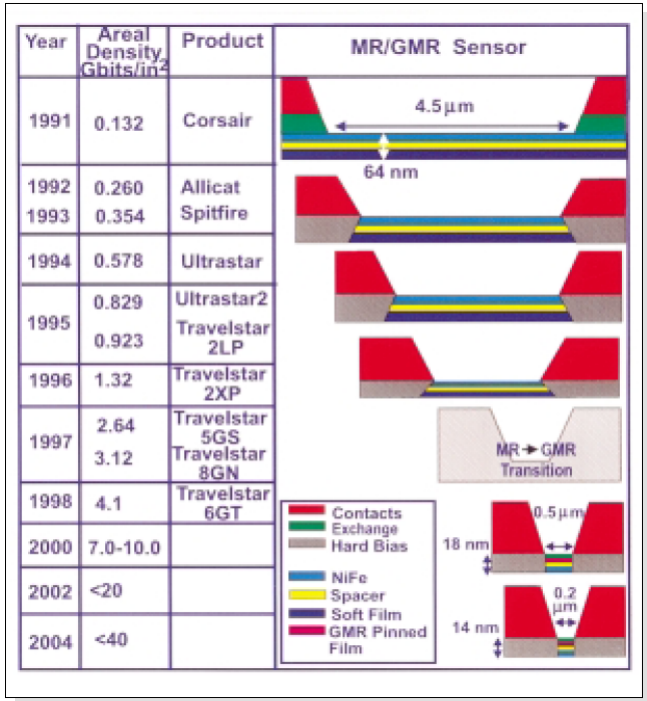
www.hgst.com

Surprisingly Accurate Dating

- Dendrochronology
- 10,000 years to the present

- Sensor Dimension Chronology
- Over a much shorter span

Year	Areal Density Gbits/in²	Product
1991	0.132	Corsair
1992	0.260	Allicat
1993	0.354	Spitfire
1994	0.578	Ultrastar
1995	0.829	Ultrastar2
	0.923	Travelstar 2LP
1996	1.32	Travelstar 2XP
1997	2.64	Travelstar 5GS
	3.12	Travelstar 8GN
1998	4.1	Travelstar 6GT
2000	7.0-10.0	
2002	<20	
2004	<40	

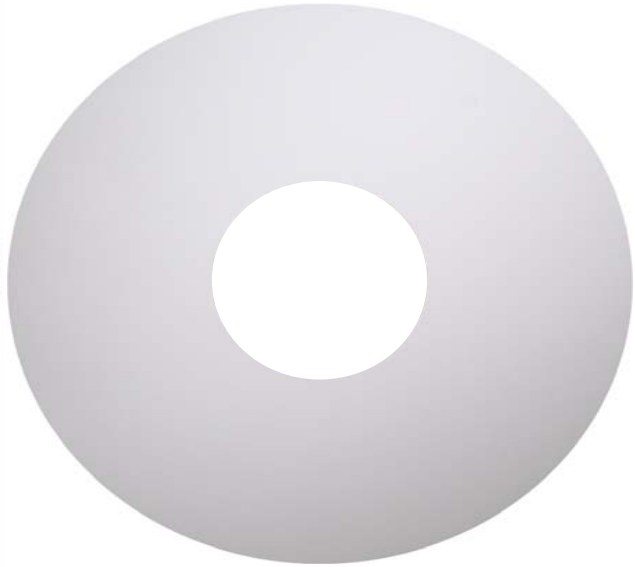


Other unique dating methods for the drive will probably also be available

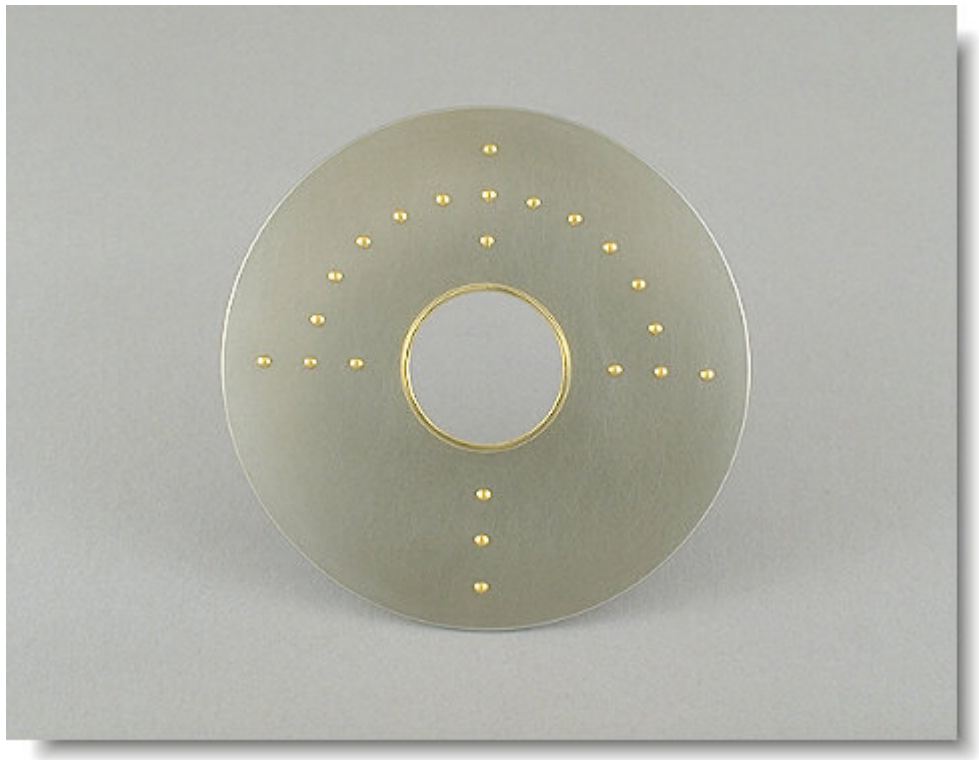
Misinterpreting the Past

Fossils and photography from the Steven Brittenham collection

Ostrich Egg Shell Bead, Phaistos Disc, Hard Disk



Perhaps it is a Brooch

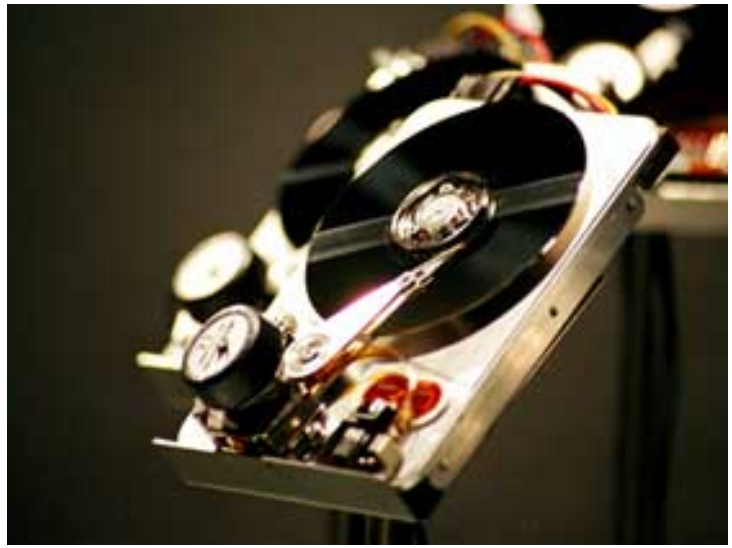


www.VelvetDivinci.com; Artist: Abrasha

Or Belt Bling?

www.HGST.com

16-Note Musical Instrument?



Valentina Vuksic's Harddisko

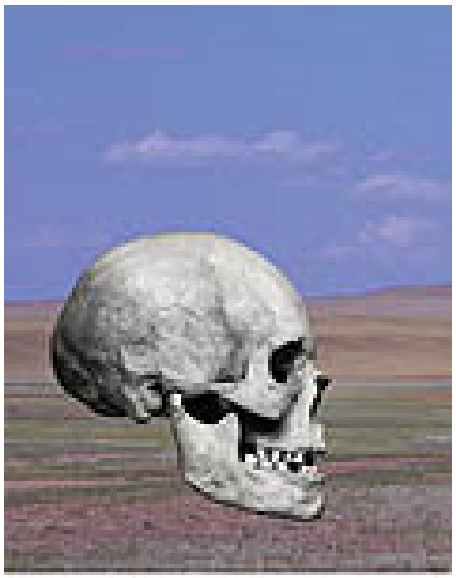
Wind Chime?

http://halogen.note.amherst.edu/~wing/wingie/tech/hdchime/hdchime.php

Furniture?

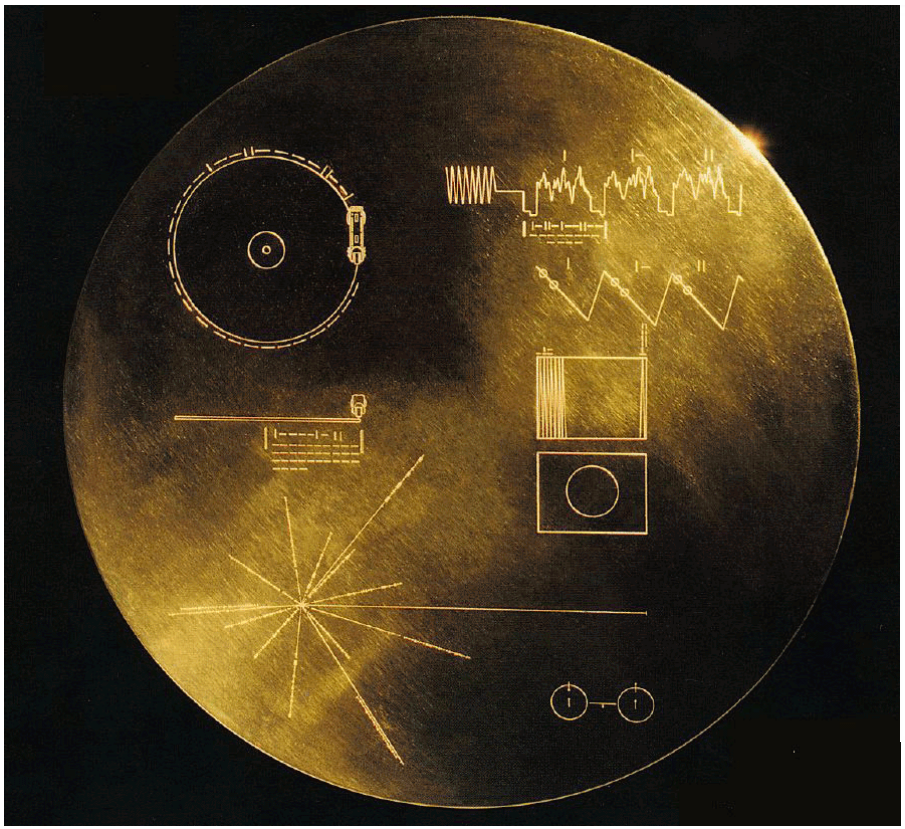
Courtesy Prof. Cai Yang, CMU
www.contrib.andrew.cmu.edu/~ycai/

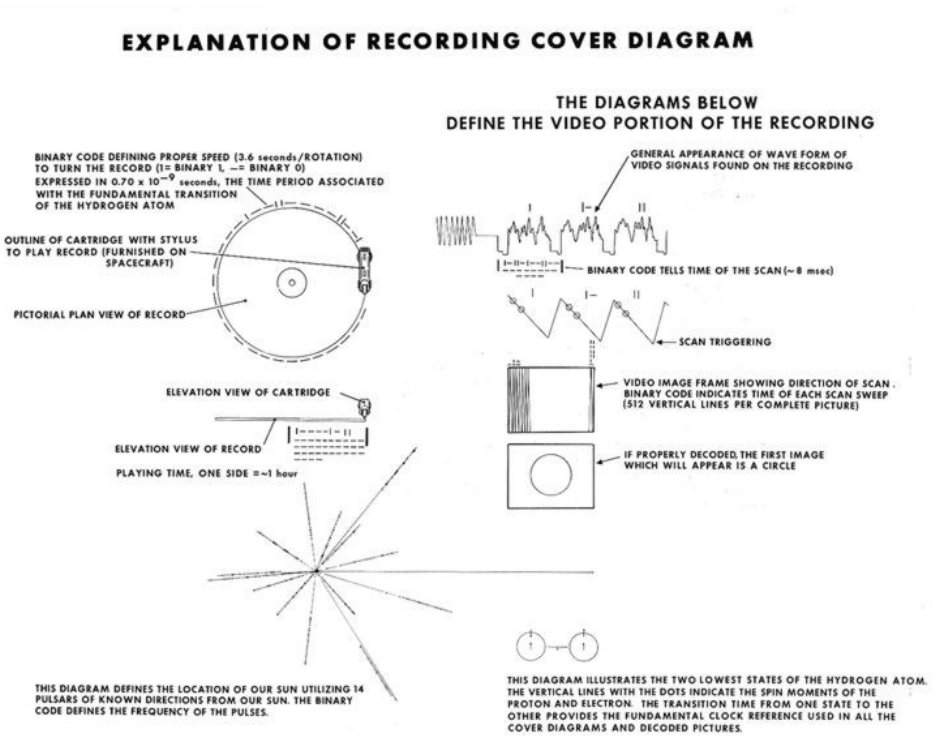
Reconstruction License: Botai-man Example



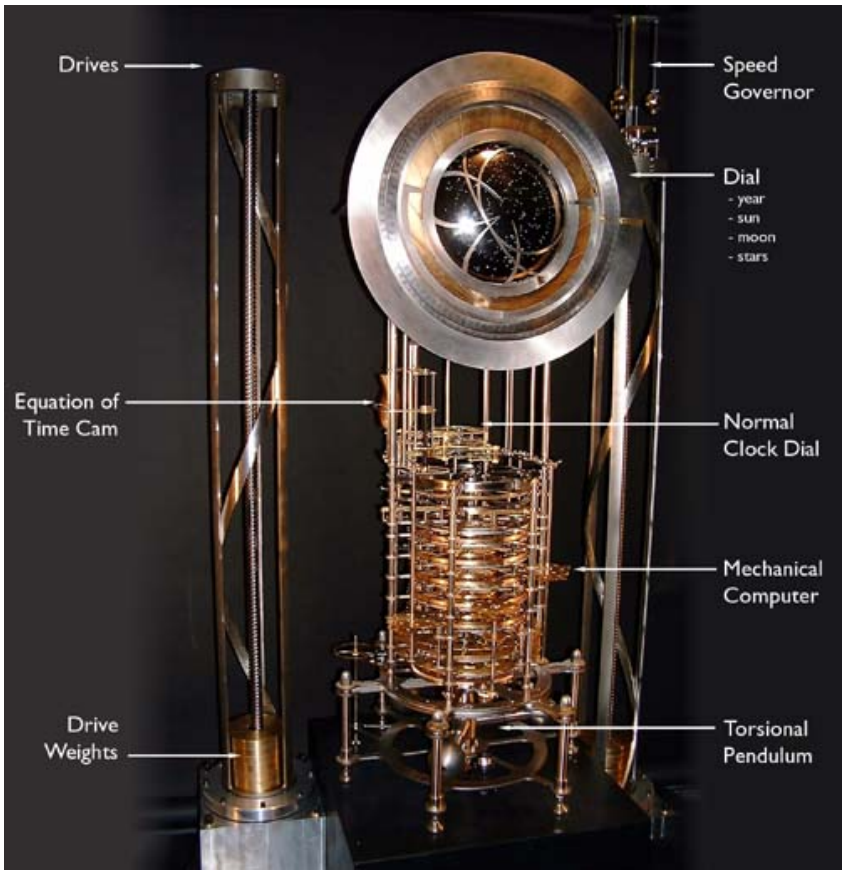
Courtesy Prof. Cai Yang, CMU
www.contrib.andrew.cmu.edu/~ycai/

Messages to the Future (> 40,000 years)
The Voyager Golden Record

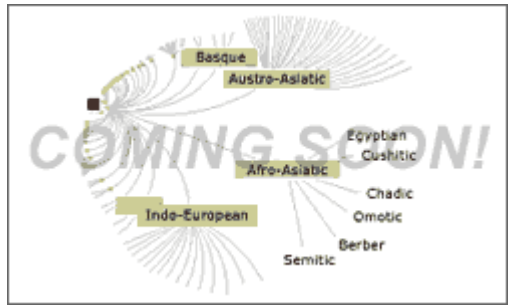




Planning for the Future: www.LongNow.com

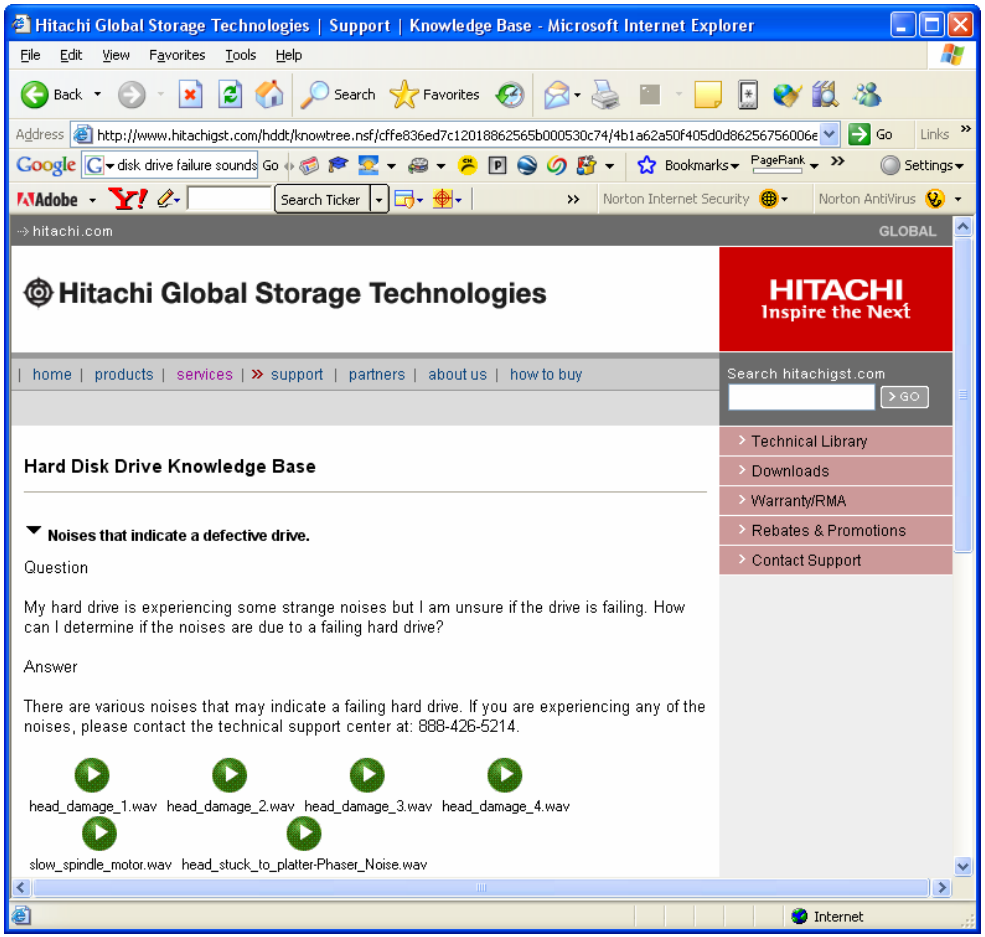


10,000 year clock



Rosetta Project
2300 languages

Did the Background Music Sound Familiar?



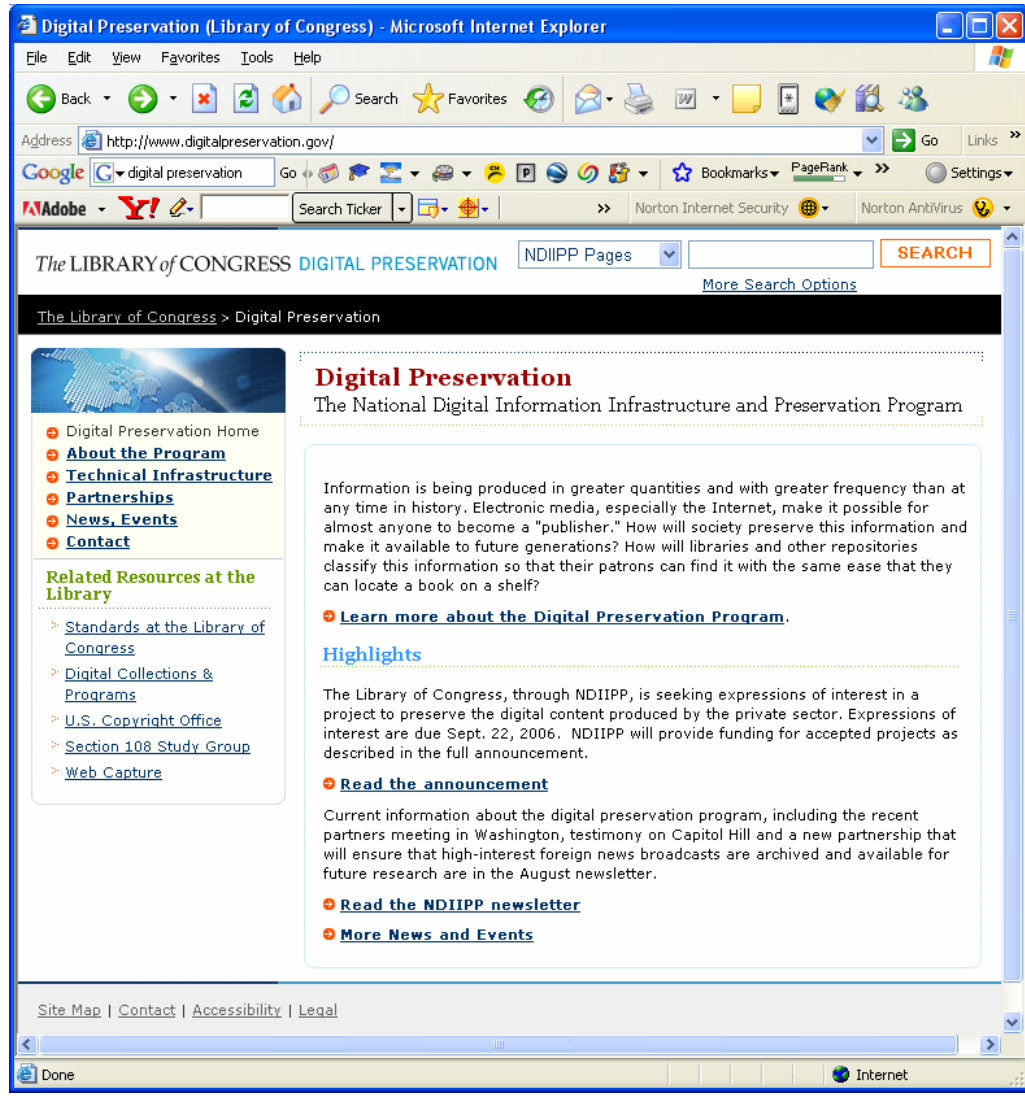
Gizmodo Hard Drive Dying Dance Track Competition Winner

James Postlethwaite
"Hitachi Hard Drive Project – Noriko Version"

www.Gizmodo.com

You may also want to hear
"My Hard Drive Crizzash"
A runner-up

Preserving Our Digital Heritage



Some Reasonable Actions are Counter-Productive to Long-term Preservation

- Efficient backups
  - Only one copy exists
- Downside of accountability legislation
  - Instead of keeping data around for discovery
  - Some data are on a 3-month deletion cycle
- Encryption
- Digital Rights Management
  - No expiration of protection?

Ready for One More Misinterpretation? Maybe They were Speakers?

www.AfroTechMods.com

Despite that kind of “Invention,” Humanity, Culture, and Science Have Endured

- In the past, data from the dominant civilization was saved by other cultures
- We have an increasingly global society
- When the world is one, there is no backup!

LOCKSS (Stanford University)

- Selected the hard disk drive as the preservation medium of choice!
- “Lots of Copies Keep Stuff Safe”
- Open source, peer-to-peer, distributed content
- Copies regularly compared
- Storage device automatically updates with user upgrade cycle

Take-aways for an HDD-related Company

- It is OK to devote resources to the very long term thinking
  - Microsoft Research – Digital Immortality
  - British Telecom – Futurology Unit
  - The very long view can spur new thinking and new connections
- Projects with a long view: Break the tyranny of areal density
  - Design an HDD to last 50 years
  - Design a drive to be built by multiple generations
    - Simulate this by multiple, staggered teams
  - Create a Voyager “Golden Disk Drive” cover explanation
  - We’re very practical!
    - Use as team building or “sabbatical” reward
    - Feed lessons-learned back into production
- The long-term future of data is uncertain
  - Where there is a need, there is a business opportunity!

What can We do Now to Improve the Future of Data?

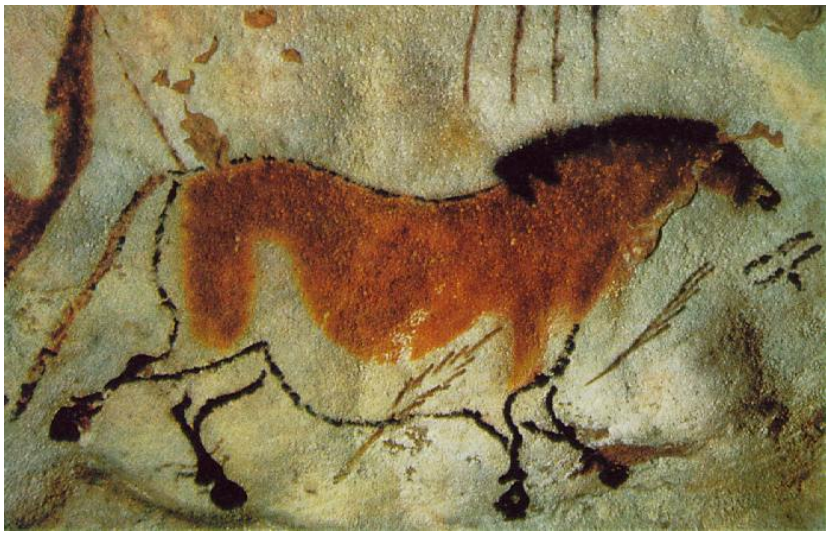
- Recoverable drives
  - Re-run drive optimization for parts replaced in the field
  - Data recovery mode
    - Simple set of universal instructions
    - IDEMA could form a committee!
- Open source
  - Encodings
  - Defect management
  - System area information
  - At least for obsolete drives!

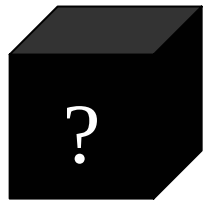
If We Don't Do It, Who Will?

Our industry is critical to the future (as well as the present)

- Think more than 5-10 years out
- Plan for our digital legacy

- Storing data is a noble profession





Thank You

LSI Logic, Inc.

For sponsoring this talk

Acknowledgement:
Thank you all for your Help!

- Barry Bailey
- Gordon Bell
- Stewart Brand
- Steve Brittenham
- Brian Fagan
- Ed Grochowski
- Bill Higgins
- Al Hoagland
- Jay Hoinville
- Robert Kemper
- Mike Mallary
- Cai Yang
- Reto Wettach
- And anyone I forgot!

In Honor of the Innovators of the Past 50 Years
Thucydides, *The Funeral Oration of Pericles*

The whole earth is the sepulcher of famous men; and their story is not graven only on stone over their native earth, but lives far away, without visible symbol, woven into the stuff of other men's lives. For you now it remains to rival what they have done.

Appears on the dedication page of the classic archeology textbook, *In the Beginning*, by Brian M. Fagan (www.BrianFagan.com)