

Losing the First Draft of History

Newspaper Content
in a Digital World

Victoria McCargar, Los Angeles Times

Opening remark

We are
doomed.

Quote

'The Nineties
are gone.'

— Julio Vera, MLIS

UCLA Dept. of Library and Information Studies,
NARA/InterPARES researcher

Quote

'Information technology is revolutionizing our concept of record keeping in an upheaval as great as the introduction of printing, if not of writing itself.'

– Jeff Rothenberg, RAND Corp.

Quote

'Are commercialization and digitizing policies depriving us of access to valuable historic images? Digital files may have only a limited lifespan of a few years before their format is obsolete, their equipment platform unobtainable, or their medium corrupted.'

— Announcement of digital archiving conference in the U.K., April, 2002

Bit Stream 101

Rothenberg, 1995

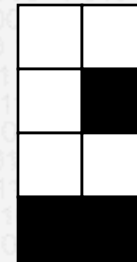
no, no, no,
yes, no, yes,
no, yes

U

21

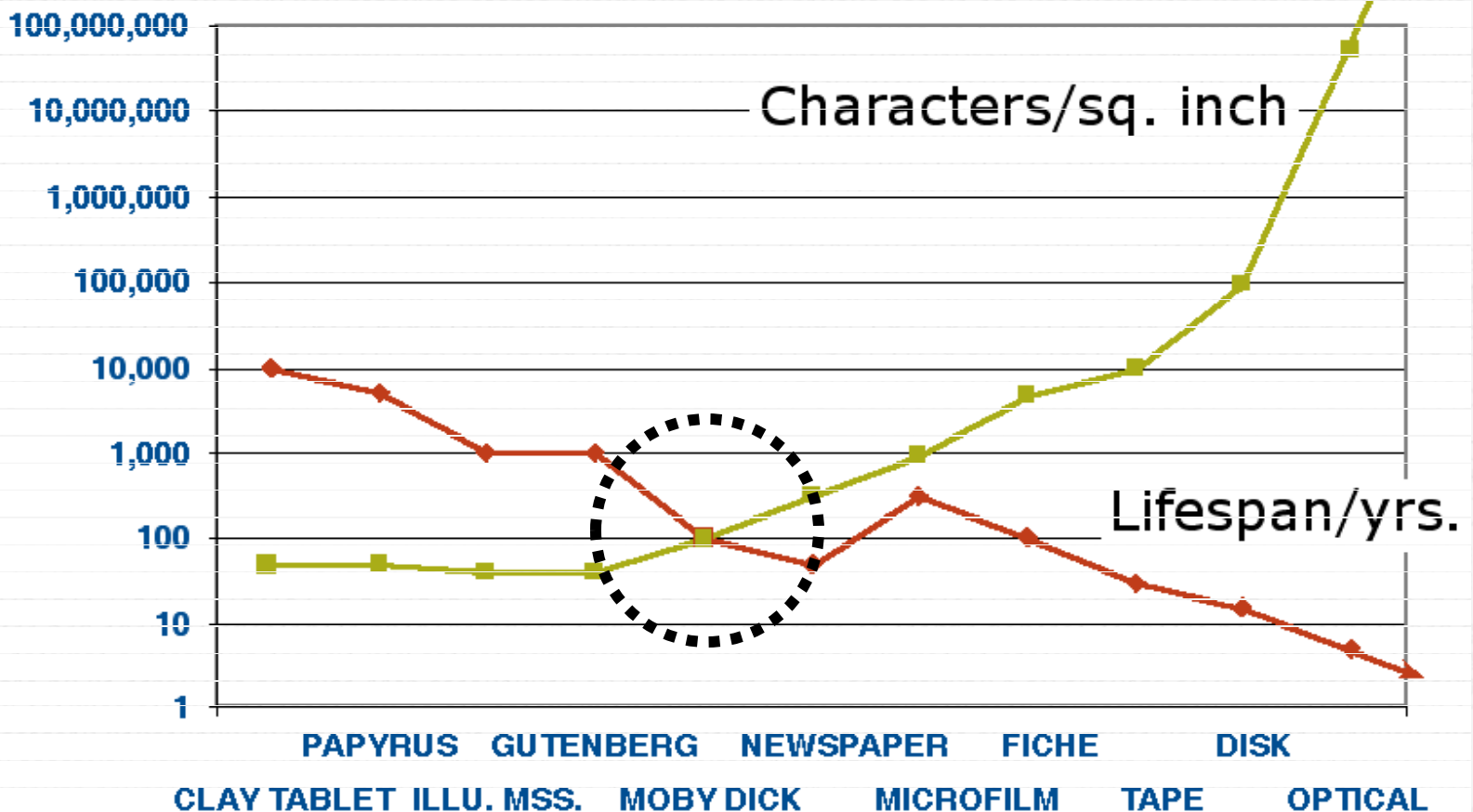
01011100001000101010000000010000

1.3125



'Irony of modern media'

Conway, 2000

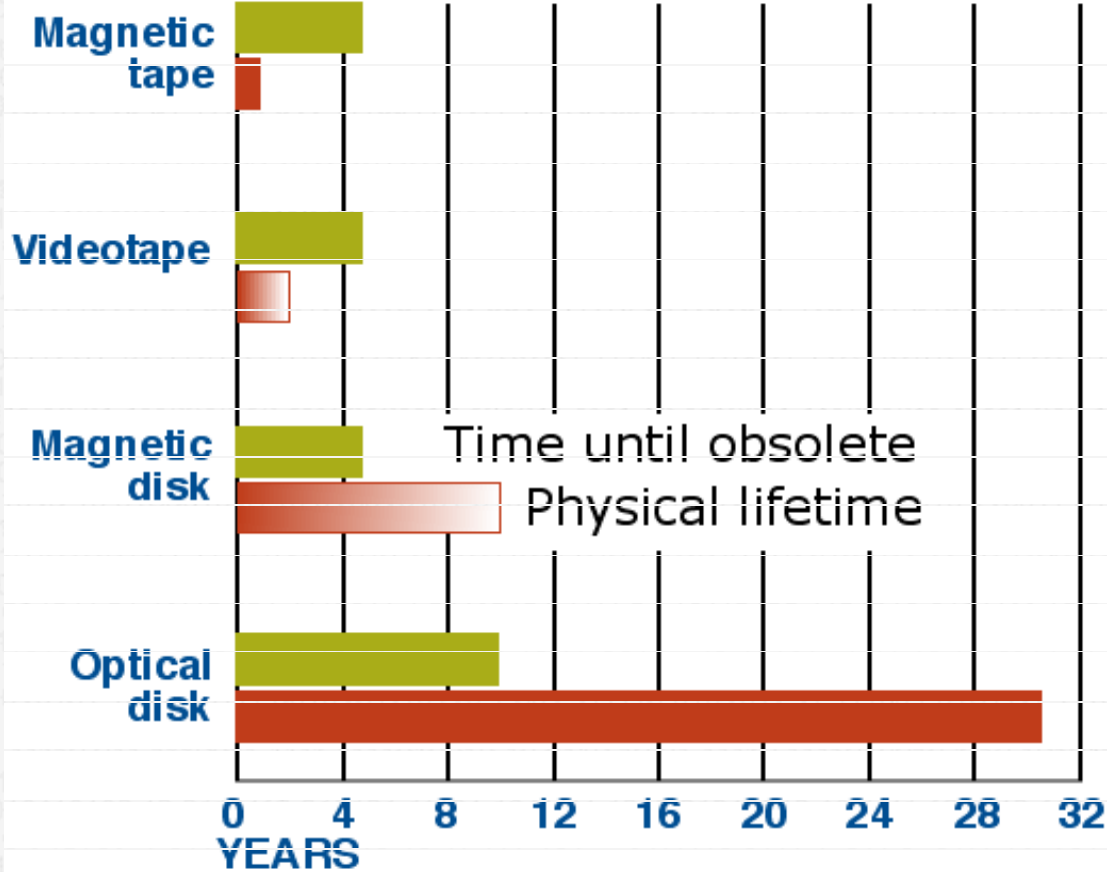


Quote

'Digital files will last forever or five years, whichever happens first.'

— Jeff Rothenberg, address to New England Library and Information Network (NELINET), Conference on Digital Preservation, June 8, 2000

Expected lifetimes



Rothenberg,
1995

Quotes

'I had no idea.'

– Former photo director of large West Coast newspaper

'I had no idea.'

– Los Angeles-based commercial photographer

'We had no idea.'

– Academic librarians at NELINET conference on digital preservation

Quotes

'I had no idea.'

—Benedictine monk in middle of project to digitize oral histories in MPEG on CD-ROM

'I had no idea.'

—Proud father of a new baby boy who just invested in an expensive digital camera

'I had no idea.'

—Friend planning to scan all her family photos

Quote

'One of the reasons people haven't figured this out yet is that there hasn't been a sufficiently large loss yet, in financial terms.'

- Anne Gilliland-Swetland, U.S. co-chair InterPARES, assistant professor, UCLA Dept. of Library and Information Studies,

Challenges I

- Organizational culture
 - Do the right people 'get it'?
 - Is the culture keeping pace with technological change?
- Technological obsolescence
 - Hardware, software, and infrastructure
 - Encryption
- Designing the right metadata schema

Challenges I

- Lack of standards
- Licensing issues
 - The content itself (stories, images, etc.)
 - *Tasini* vs. 'preservation copies'
 - Copyright may be an even worse problem than technology
- Economic issues
 - How much will it cost? No one knows
- Education

Quote

'The granddaddy of [all digital records projects] has an announcement on their web site today saying their records were all accidentally deleted by their systems people.'

— Anne Gilliland-Swetland, April 8, 2002

Research trends I

- Emulation: Jeff Rothenberg
 - Archiving information about the bit stream along with the information
 - Program computers to behave like older machines
 - Addresses *behavior*
 - Still theoretical, questions about scalability

Research trends I I

- More emulation
 - Once and for all
 - Fixed functionality
 - Fixed data
 - Certainty
 - Independent of value and use

Research trends III

- Migration (more real world)
 - Requires periodic intervention
 - Accounts for evolving functionality
 - Possible corruption
 - Uncertainty
 - Dependent on value and use

Sidebar

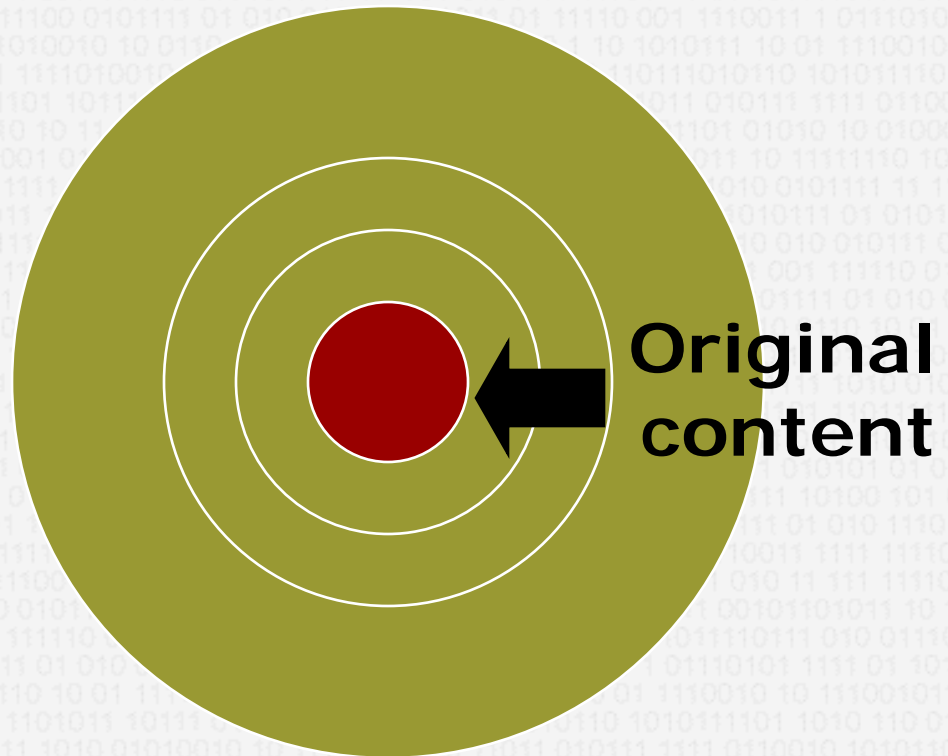
Migration in action



The metadata spiral

- Migration 1
- Migration 2
- Migration 3

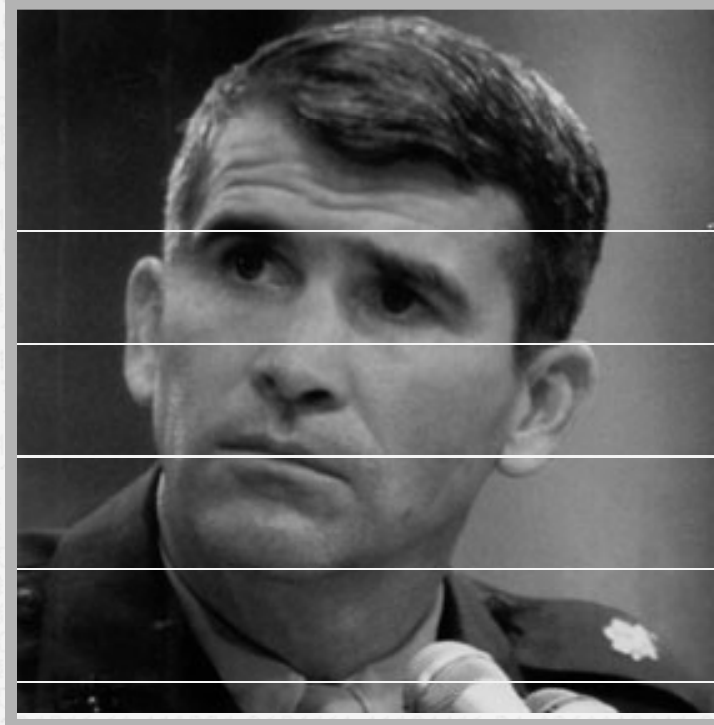
You're now migrating more metadata than content.



What do they have in common?



San Diego
Supercomputer
Center



Lt. Col. Oliver North

'Persistent objects'



- SDSC loves a challenge
- 'Decomposing' software components and reassembling them
- XML strategy
- It's only e-mail!

Research trends I V

- OAIS (Open Archival Information System)
 - NASA, NARA, InterPARES, many others
 - Deep description of data that aids in reassembly at the point of query
 - Addresses issue of complex, distributed objects; less about behavior

Critical concepts

- It is not possible to preserve an electronic record—you can only preserve the ability to create a new version when you need to
- To do this, you need a lot of metadata
- One file may really be multiple files
- How carefully you go about this depends on how willing you are to risk losing images
- Cataloging is not dead

What it will take

- Intensity of warrant or risk (how badly you need or want to do it)
- Staff with a lot of clout
- Mature systems
- Organizational culture
- Standards, standards, standards

Standards

- Industry 'standards' (PDF, JPEG, EPS) are risky
 - *De facto vs. de jure* standards
- 'The great thing about standards is that there are so many of them.'
- Easier to hack
- Lowest common denominator

Quote

'Standards are never cutting edge, by definition.'

— Anne Gilliland-Swetland

Hard questions

- Who is going to do this?
- Where and how will we do it?
- How much will it cost?
- How do we monitor it year after year?
 - Generation after generation?
- What about copyright?
- Where are our standards going to come from?

System criteria

- Quality of what goes in (content, metadata)
- Who puts them in (do they know what they're doing)
- How are images removed and by whom (you can't keep everything forever)
 - *Choose what to lose*

Quote

'Now that we're making all these PDFs of pages, why can't we stop doing microfilm?'

— Former production superintendent

What can we do?

- Educate yourself
- Keep up with trends
- Keep microfilming
- Maintain good cataloging and indexing practices
- Ask vendors what their strategies are
- Help get the word out

Quotes

'Libraries have to figure out preservation... We're counting on you to do it.□'

— Tim Berners-Lee, MIT, inventor of the World Wide Web, address to NELINET preservation conference

'[Digital] preservation is a job for libraries.'

— Paul Conway, Yale University

Quote

**BOK!
BOK!
BOK! BOK!
BOK! BOK! BOK!
BOK!**

—Chicken Little