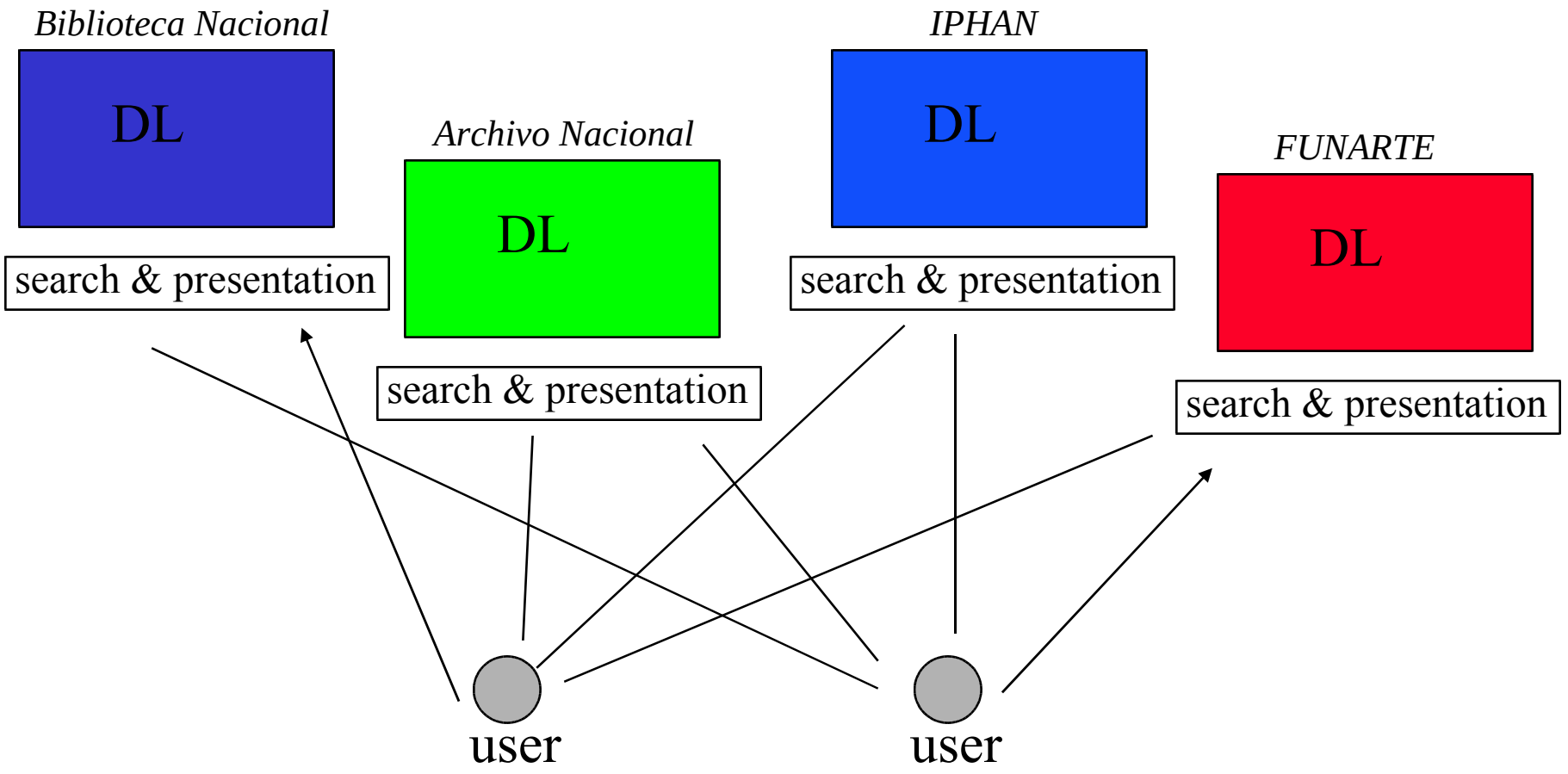
# Challenges for Digital Preservation: Standards, Architectures, & Copyright

Howard Besser
New York University
Moving Image Archiving & Preservation Program
http://besser.tsoa.nyu.edu/howard

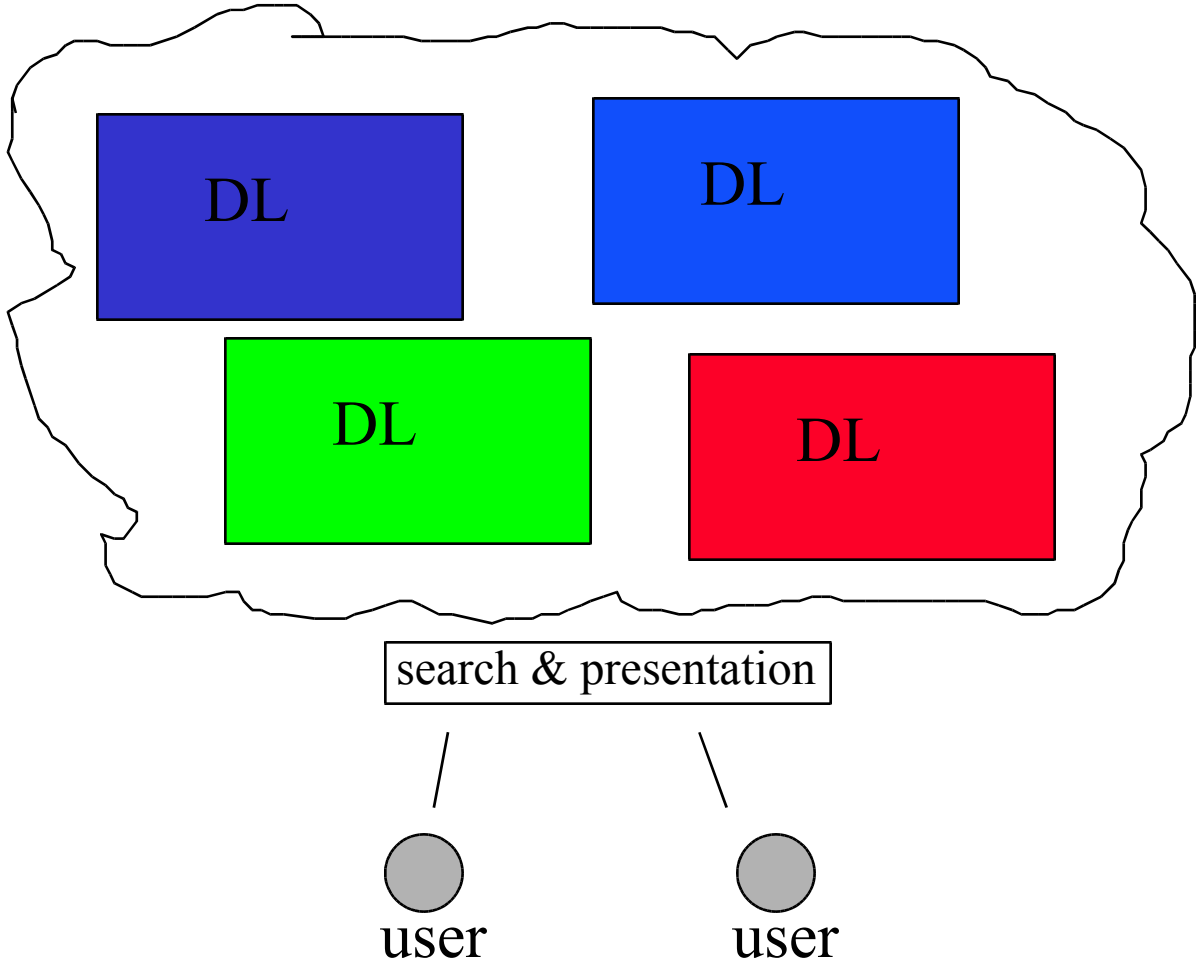
# Challenges for Digital Preservation: Standards, Architectures, & Copyright

- Models for Digital Repositories
- The importance of Standards
- Special problems with Digital Preservation
- Paradigm shift for Preservation
- Basic Standards issues
  - Best Practices for Image capture & storage formats
  - Longevity Issues & Preservation Repositories
  - Other types of metadata standards (Image Identification, Standards & other Metadata, Best Practices, …)
- Digital Preservation activities
- The horrors of Copyright changes

# Traditional Digital Repository Model



# Ideal Digital Repository Model



# Resources that use elements of the Ideal Model

- Open Archives Initiative (OAI) and metadata harvesting
- Electronic Theses & Dissertations
- Budapest Open Access

# From Digital Collections to Digital Libraries, Museums, and Archives

- No longer merely experiments
- Adhere to our fields' traditions (access, interoperability, sustainability, privacy, …)
- Provide services

# For Interoperability, Repositories Need Standards

(as well as Sustainability & Access)

- Technical standards for architecture
- Descriptive Metadata for consistent description
- Discovery Metadata for finding
- Administrative Metadata for viewing and maintaining
- Structural Metadata for navigation
- ... Terms & Conditions Metadata for controlling access...

# Why are Standards and Metadata consensus important?

- Managing digital files over time
- Longevity
- Interoperability
- Veracity
- Recording in a consistent manner
- Will give vendors incentive to create applications that support this

# How long we keep things

- Companies keep information for days, or even years
- Individuals keep things for years, or a lifetime
- Archives, Libraries, and museums keep things for hundreds of years

*Cultural Institutions have a much greater responsibility for preservation!*

# But who is preserving today's "*born-digital*" works?

- **In the past, we knew about history by finding written documents:**
  - Changes between different drafts of a scientific or literary paper
  - Letters and correspondence between a scientist (or literary figure) and colleagues (that both helps contextualize the work, and lets us see changes in thought processes or discovery)
- **But today, these documents are not on paper! They are in the form of:**
  - Email correspondence
  - Word processing files that do not show changes between drafts/versions
- **Who will take responsibility to save these works for future study?**

# Serious Longevity Problems

- Previous formats required little ongoing intervention (remote storage facilities, Iron Mtn); digital formats require intense ongoing management
- The Short Life of  Digital Info-

# The Short Life of Digital Info: Digital Longevity Problems

- Disappearing Information
- The Viewing Problem
- The Scrambling Problem
- The Inter-relation Problem
- The Custodial Problem
- The Translation Problem

# The Viewing Problem

- Digital Info requires a whole infrastructure to view it
- Each piece of that infrastructure is changing at an incredibly rapid rate
- How can we ever hope to deal with all the permutations and combinations

# The Scrambling Problem

Dangers from:

- Compression to ease storage & delivery
- Container Architecture to enhance digital commerce

# The Inter-relation Problem

- -Info is increasingly inter-related to other info
- -How do we make our own Info persist when it points to and integrates with Info owned by others?
- -What is the boundary of a set of information (or even of a digital object)?

# The Custodial Problem

- In the past, much of survival was due to redundancy
- How do we decide what to save?
- Who should save it?
  - Mellon-funded E-Journal Archives
- How should they save it?-

# The Custodial Problem:
## How to save information?

- Methods for later access
  - Refreshing
  - Migration
  - Emulation
- Issues of authenticity and evidence

# The Translation Problem

- Content translated into new delivery devices changes meaning
  - -A photo vs. a painting
  - -If Info is produced originally in digital form in one encoded format, will it be the same when translated into another format?
  - Behaviors

# Digital Images are very fragile; We need to be even more careful about:

- Capture (image quality)
- Storage (file format)
- Management (refreshing, migration/emulation, linking metadata to “essence”)

# Paradigm shift

- From preserving a physical artifact to
- Preserving disembodied digital content

- Instead of primarily worrying about atmospheric control
- Worry about constant management of the digital works

# Paradigms Shifts needed

| | **Old** | **New** |
|---|---|---|
| **Physical preservation** | atmospheric cntrl | ongoing mgmt |
| **What to save?** | artifact | idea + ancillary material & documentation |
| **Cataloging** | Individual work in hand | FRBR |
| **Later access** | Artifact & documentation | Restaging, ancillary material & documentation |

# Risk Management

- We can’t say definitively that we can make every digital work persist
- What we CAN say is that the more a digital work conforms to standards and best practices, the greater the likelihood that we can assure persistance
- Our preservation repositories can even accept deposits of non-conforming works, but the less they conform, the less likely that they’ll be salvageable
- Persistance is most likely for works that share standards, metadata, and best practices

Capture & Storage:

# Reformatting Best Practices

- Use/Users/Collection
- Benchmarking
- Masters vs. Derivatives
- Scanning-
- Administrative Metadata-
- Structural Metadata-

# CDL/TASD Standards & Best Practices
http://www.cdlib.org/about/publications/

- *California Digital Library Digital Object Standard: Metadata, Content and Encoding*, May 18, 2001
- *California Digital Library Digital Image Format Standards*, July 9, 2001
  http://www.cdlib.org/about/publications/

# Scanning Best Practices

- Think about users (and potential users), uses, and type of material/collection
- Scan at the highest quality that does not exceed the likely potential users/uses/material
- Do not let today's delivery limitations influence your scanning file sizes; understand the difference between digital masters and derivative files used for delivery
- Many documents which appear to be bitonal actually are better represented with greyscale scans
- Include color bar and ruler in the scan
- Use objective measurements to determine scanner settings (do NOT attempt to make the image good on your particular monitor or use image processing to color correct)
- Don't use lossy compression
- Store in a common (standardized) file format
- Capture as much metadata as is reasonably possiple (including metadata about the scanning process itself)

To deal with variant forms:

# Incorporate parts of Functional Requirements for Bibliographic Records (FRBR)

- work
- expression
- manifestion
- item

# NISO/DLF Technical Image Metadata Workshop--4/99

(Z39.87-2002 draft)

- create metadata needed to manage images in digital repositories over long periods of time (full life-cycle mgmt)
- document image provenance & history
- ensure that the images will be rendered accurately on any output device

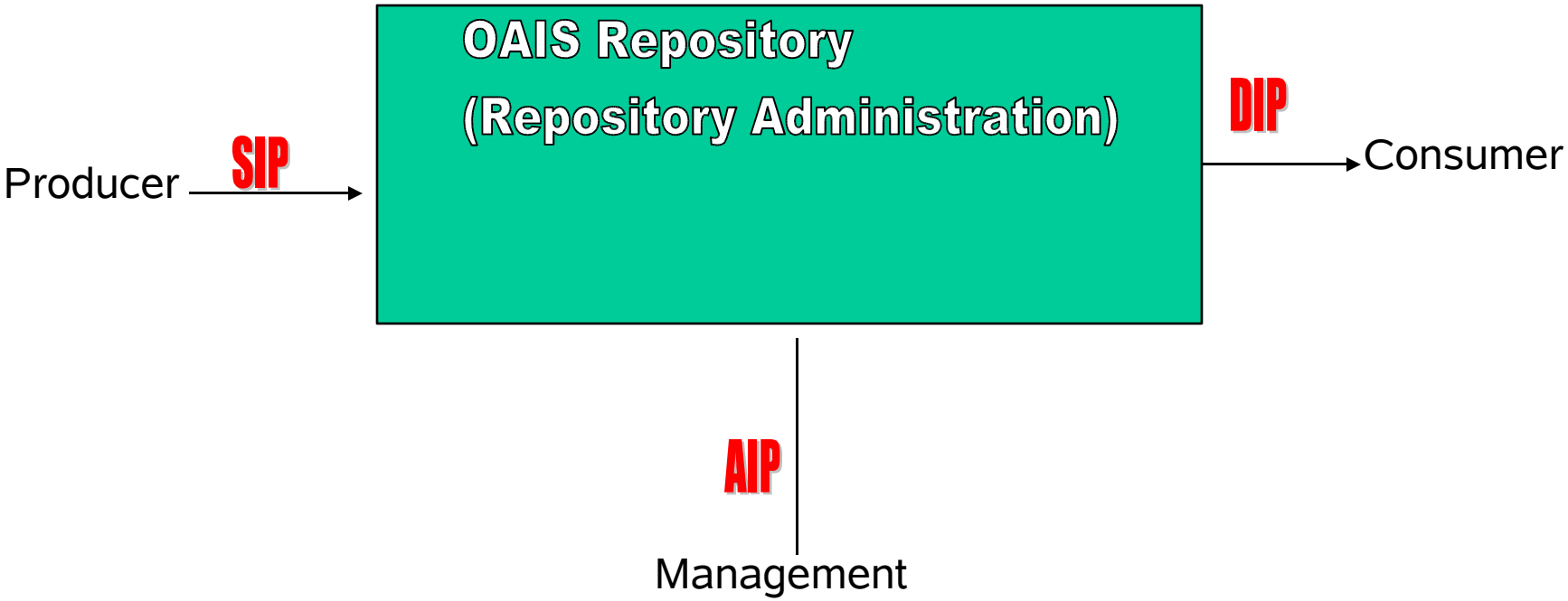
# Technical Image Metadata-Z39.87

- Image parameters (MIME type, compression, colorspace & profile, …)
- Image Creation (source, capture info, etc.)
- Image performance assessment (sampling, colormap, whitepoint, target data, etc.)
- Change history (source, processing, etc.)

- additional XML implementation schema (MIX)

Management:

# Preservation Repositories: Open Archival Info System Model



# Preservation Repositories: Open Archival Info System Model

- High-level reference model describing submission, organization and management, and continuing access
- Conceptual framework for different organizations to share discussions with a common language
- Producers, consumers, management, actual repository
- SIP, DIP, AIP
- AIP consists of data objects plus representation info (Content, Preservation Description, Packaging, Descriptive)
- Originally developed for Space Science community

# OCLC/RLG Digital Repository Attributes

- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological suitability
- System security
- Procedural accountability

# OCLC/RLG
# Selected Recommendations

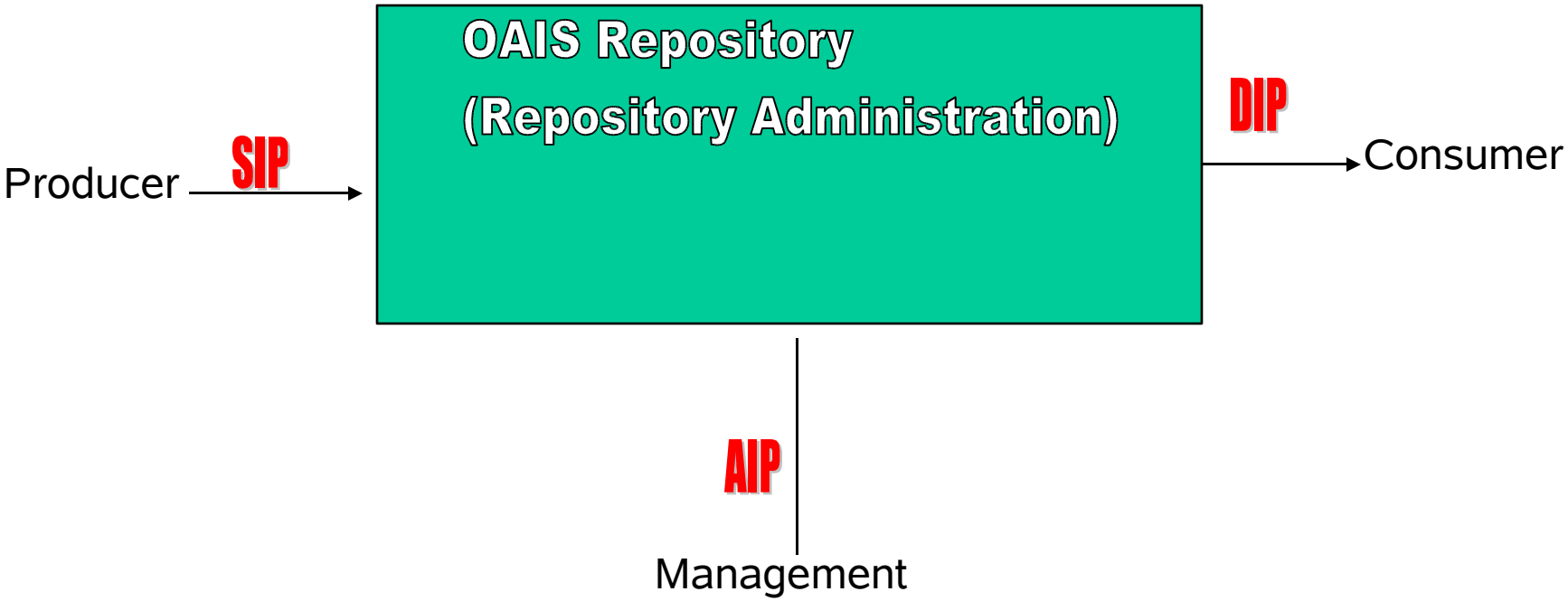
- Policies, Certification processes, Risk management, Persistent ID, Migration/Emulation experiments
- Stakeholders meet to decide how to describe what is in a dig repository
- Examine special properties of particular classes of digital objects
- Technical standards for exchange and interoperability btwn repositories
- Develop projects and case studies
- Copyright issues

# Preservation Repositories:
## too difficult for small institutions

- Too complex for small institutions to manage
- Will be done through partnering (small museum with University) or through consortia (museum association, state-wide organization, …)
- Archive or museum will direct what is needed, but digital repository will carry out the actual work (as defined in SIP/DIP/AIP)

Management:

# Preservation Repositories: Open Archival Info System Model



# Preservation Repositories: Projects based on OAIS Model

- CEDARS
- NEDLIB
- Pandora
- CDL
- OCLC/RLG Working Group on Preservation Metadata, *Attributes of a Trusted Digital Repository*, August 2001-

# Preservation Metadata

- OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata for Digital Objects: A Review of the State of the Art*, January 31 2001
- OCLC/RLG Working Group on Preservation Metadata, *A Recommendation for Content Information*, October 2001

# Preservation Repositories -- AIP Metadata

- Preservation Description Info
  - reference info
  - context info
  - provenance info
  - fixity info
- Packaging Info
- Descriptive Info
- Content Info

OCLC/RLG Efforts

# Working Group I: Preservation Metadata Framework

- …to define the concept of preservation metadata, describe its importance in context of the overall digital preservation process, examine the "state-of the-art" in the use of metadata in support of digital preservation, and evaluate the prospects for a community-wide, consensus-building activity in the area of preservation metadata (Preservation Metadata for Digital Objects: A Review of the State of the Art http://www.oclc.org/research/pmwg/presmeta_wp.pdf)
- …to develop a framework outlining the types of information—i.e., metadata—that should be associated with an archived digital object. (A Metadata Framework to Support the Preservation of Digital Objects http://www.oclc.org/research/pmwg/pm_framework.pdf)
  - an expanded conceptual structure for the Open Archival Information System (OAIS) information model, and
  - a set of metadata elements, mapped to the conceptual structure and reflecting the information concepts and requirements articulated in the OAIS model.

OCLC/RLG Efforts

# Working Group II: PREservation Implementation Strategies (PREMIS)

- develop a core set of implementable preservation metadata elements, with broad applicability within the digital preservation community
- develop a data dictionary to support the preservation metadata element set
- examine and evaluate alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system, as well as for the exchange of preservation metadata between systems
- develop a pilot program for testing the group's recommendations and best practices in a variety of systems settings
- explore opportunities for the cooperative creation and sharing of preservation metadata

# Structural & Administrative Metadata (METS)

- For complex, multi-part works (collections of images, groups of maps, albums of photographs, etc.)
- Structural--preserving relationships (books/chapters/sections/pages, photo albums on a page, master images with thumbnails and different sizes, etc.)
- Administrative--keeping information about managing the works over time (keeping track of all of the pieces in a group, what software is needed to decompress/view an image, etc.)
- Metadata Encoding & Transmission Standards (METS)--wrap the image(s) in XML encoding

# Older Longevity Projects

http://sunsite.berkeley.edu/Longevity/

- CPA Task Force
- Getty “Time & Bits” Conference & Follow-ups-
- Preservation experiments in US and Europe
  - NEDLIB, CURL, Michigan
- Internet Archive
- Long Now

# Recent Digital Preservation Activities-

- LC Natl Dig Info Infrastructure & Preservation
- InterPARES
- Emulation Projects
- E-Journal Archiving
- ERPANET
- Persistent Naming

# LC's National Digital Information Infrastructure and Preservation Program (NDIIP)

- Authorized Dec 2000
- LC, Dept of Commerce, NARA, White House Office of Sci & Tech Policy
- with help from CLIR, NLM, NAL, OCLC, RLG
- Ongoing collab process
- Commissioned papers on preserving: the Web, periodicals, digital sound, E-Books, Digital TV, Digital Video
- Accepting applications for "Building a Network of Partners" phase (up to $3 million)

# CLIR/NDIIP Issue Areas

- Technical and architectural infrastructure (standards, ID, obsolescence)
- Economic and legal (rights mgmt, funding)
- Collection Development (what gets saved?)
- Societal & Institutional (who does what, role for commercial sector)

# Research: NYU/Public Television Project

- Preserve a broad set of elements (including ancillary material)
- Life-cycle mgmt (add metadata as soon as a clip comes in)
- Establish a community of stakeholders, working together for preservation (stations, university, librarians, journalists, historians, producers, scholars, …)
- Build an OAIS Server
- Explore appropriate file formats, wrappers, METS extensions
- Develop sustainable business model

# InterPARES 2

International Research on Permanent Authentication Records in Electronic Systems

- Ongoing international archival world project examining how to make electronically-generated records last over time
- Developing the theoretical and methodological knowledge needed, then will formulate model policies, strategies, and standards
- Reliability, accuracy, authenticity
- In 2003 was extended to include dynamic, interactive, and experiential works

# Emulation Projects

- CAMiLEON (Michigan/Leeds)
- NEDLIB

# E-Journal Archiving

- Issues
  - License, don't own; may not be even able to obtain right to make archival copy
  - Increasingly no paper back-up at all
  - Usually we don't have the important redundancy factor
- Mellon funded projects (2001)
  - Yale, Harvard, Penn working w/individual publishers
  - Cornell, NYPL--specific disciplines
  - MIT exploring characteristics that change (dynamic)\
  - Stanford--archiving software tools

# Electronic Resource Preservation and Access NETwork (ERPANET)

- Best practices and skills development for digital preservation of cultural heritage and scientific objects
- 3 year project launched Nov 2001; 1.2 million Euros

# Persistent Naming

- URNs
- Handles
- PURLs
- Re-directs

# Copyright Implications for Preservation-

- Changing Copyright environment tries to restrict access
- The “refreshing” problem
- The “emulation” problem
- Problems of access & commodification of information

# Changing Copyright environment attempts to restrict access

- World Intellectual Property Organization (WIPO) attempts to harmonize copyright legislation to benefit international commerce (large rights-holders)
  - Makes all of us live by the rules created to benefit commerce
- Legislative and judicial attempts to limit or eliminate tools of vital importance to scholars (public domain, fair use, first sale), and to dismantle our important “information commons”
- Technological attempts by rightsholders to control downstream use
- Licensing replacing sales, pay-per-use models
- Commodification of information

# Copyright Changes Threaten Scientific Research

- Scientific & technological progress builds upon the work of previous research. Restricting access to previous research is a blow to scientific development
- Using © to privatize information or to prevent criticism
  - Pay to read articles
  - Can't quote without permission
  - Can't incorporate old ideas into new (articles or inventions/patents)
  - Can't teach w/o violating © laws (online learning, student readings)
  - Entire science/technology fields eliminated (prohibition of reverse-engineering, decryption, …)
  - Threats of lawsuits have "chilling effect" (Felton case)
  - US Election machines (Diebolt) © lawsuits against people who create online exposés of the machine problems

# The “refreshing” problem

- With the need to “refresh” at least once/decade, our repositories will constantly need to make copies
- Yet, because of fear of rights-holders (piracy & control), copyright laws are being re-written to prevent our repositories from making copies

# The “Emulation” problem

- In order to emulate the applications software to view current works, we will need to preserve copies of older software versions, and sometimes reverse-engineer these
- New copyright laws are trying to make both copying software onto new strata and reverse-engineering illegal

# A few examples of copyright issues-

# Fair Use is Disappearing

- Criminalizing Fair Use
- Copyrighting DB contents in perpetituity
- Proposed Legislation (USCITA, etc.)
- The 1998 Digital Millenium Copyright Act
- The 1998 Sonny Bono Term Extension Act
- … Hollings Bill, …

# Difficulty of clearing rights affects Completeness of Content

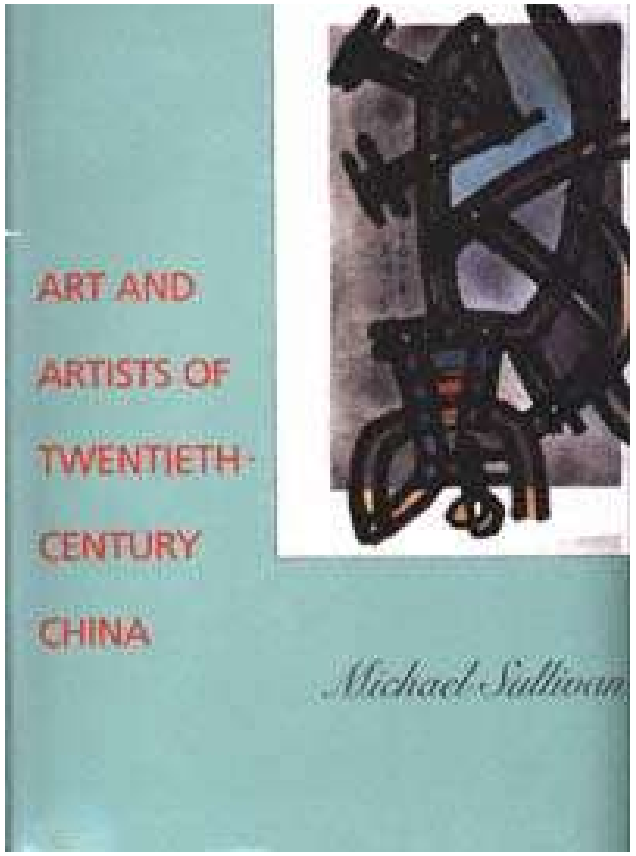
Image Not Available

Image Not Available

Image Not Available

Image Not Available

Image Not Available



# Recent changes are reducing the tools artists used to create new works based on old

- Lengthening copyright terms
- Shrinking public domain
- Technical protections that prevent one from accessing content that one may legally be able to fairly use-

# Time before works enter public domain

| **Law** | **Duration** |
| --- | --- |
| 1709 British | 14 years |
| 1790 US | 14+14 years renewal |
| 1909 US | 28+28 yearsrenewal |
| 1976 US | 75 years (corporate)<br>life+50 years (individual) |
| 1998 US | 95 years (corporate)<br>life+70 years (individual) |

# Lengthier Copyright Means

| **Date** | **Term** |
|---|---|
| Published 1923-63 | 67 years if renewed |
| Published 1964-77 | 28+67 years |
| Created before 1/1/78 | Life+70 years or<br>-12/31/02 if not published<br>-12/31/47 if published before end of 2002<br>whichever is greater |
| Created after 1/1/78 | Life+70 years (95/120 years corporate) |

# Works that should have already entered the Public Domain (but didn't)

- Virginia Woolf:  Jacob's Room
- Film -- Sherlock Jr.
- F. Scott Fitzgerald:  Hot and Cold Blood *and* Invasion of the Sanctuary
- Zane Grey:  Code of the West, Steelhead, Tappan's Burro,The Vanishing American, *and* Down into the Desert
- Ben Hecht:  Fingers at the Window
- Rudyard Kipling:  Independence *and* London Stone
- P.G. Wodehouse: Jeeves, First Aid for Dora Heart of a Goof, Leave It to Psmith, Magic Plus Fours, No Wedding Bells for Him,The Return Of Battling Billson, Rollo Podmarsh Comes To, Ukridge Rounds A Nasty Corner, *and* Chester Forgets Himself

# Works that Sonny Bono prevented from entering the Public Domain

- Irving Berlin: ***Blue Skies*** (2002)
- Harry Woods: ***When the Red, Red Robin Comes Bob, Bob Bobbin' Along*** (2002)
- Oscar Hammerstein II and Jerome Kern: ***Ol' Man River*** and ***Showboat*** (2003)
- Mickey Mouse (2004)

# Copyrighting Facts: Proposed Database Extraction legislation

- No requirement that the DB contain any original content (can copyright facts, government information, etc., taking these out of the Public Domain)
- DB owner given recourse, even if they didn't suffer harm
- Implications on:
  - Transformative uses
  - Uses other than those intended by the compiler (citation analysis)
  - Copyrighting court decisions

# Use of IP Laws to inhibit free speech and stifle creativity

- The Wind Done Gone
- E-Toy
- Leonardo Finance vs Leonardo Arts
- Jeff Koons case
- Barbie
- Scientology vs. Netcom
- Fans sites (Star Trek, Harry Potter)
- 2 Live Crew
- Negativland and U2
- Contract with America
- Snow White & AIDS
- Disney and Dan O'Neil
- The Rio player

# Like Scientific Info, Art & Popular Culture Images are part of our heritage

- Raw material for public discourse
- Appropriation is important for a democracy and a Civil Society

Historically, copying of Art was viewed as ***homage***, or at least as perfectly acceptable commentary

*Christ Carrying the Cross*
Thomas de Coloswar
Tempera on pine, 1427

*Christ Carrying the Cross*
Hieronymus Bosch
Oil on Panel, 1485-1490

*Christ Carrying the Cross*
Hans Multscher, 1437

*Christ Carrying the Cross*
Cornelis Floris
Marble, c.1570

Besser-ISTEC, 21/5/04

# Madonna & Child

*The Adoration, with the Infant Baptist and St. Bernard*
Fra Filippo Lippi
Tempera on panel, c.1432

*Wilderness Adoration of the Child with Saint John*
Jacopo di Poggibonsi
Tempera on panel, c.1443-45.

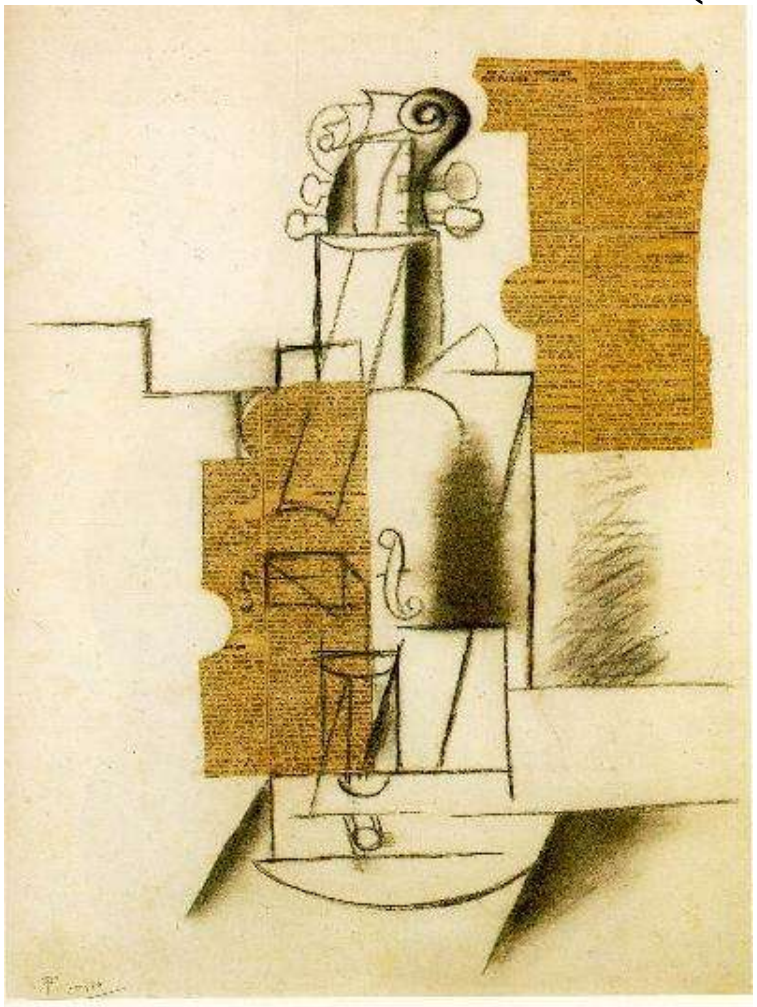
*Adoration of the Child with Saint John*
Jacopo di Poggibonsi
Tempera on panel, c.1443-45

# Art (particularly 20th century) builds upon prior Works

new works involve repurposing old

- Collage
- jazz
- Pop
- Postmodern
- Rap/Sampling
- Shakespeare

# Picasso's *Violin* (1912)



# Duchamp (1919)

# Kruger's **Untitled 1990 (It's a small world but not if you have to clean it)**


underlying image from Thomas Hoepker's "Charlotte as seen by Thomas,"
Originally published in German photography magazine *Foto Prisma* in 1960

# Strong © protection/DRM can inhibit Teaching

- Sound or film clips
- dist learn violating ©-
- Teach Act
- My stories
- Threats for posting readings
- Curricular Materials

# Distance Learning: Violating ©

# Importance of Info Commons for Content

- Common Heritage (philosophical)
- New Knowledge incorporates Old (progress)
- Derivative Works rely upon pre-existing Works (creativity)
- Social Commentary (free speech)

# Dangers

- Eliminating a public domain of information
- Controlling social/political commentary, satire, creation of  new derivative works/recombinant works
- Criminalizing acts that might possibly impede digital commerce
- Making sure that the Internet is used only for info consumption, not production
- Controlling access to older info (controlling history)

# Public Space is becoming Private

Private Property. Permission to pass revocable at any time.

# The new paradigm: Ongoing Management

- Previous formats required little ongoing intervention (remote storage facilities, Iron Mtn); digital formats require intense ongoing management

# Paradigms Shifts needed

| | **Old** | **New** |
|---|---|---|
| **Physical preservation** | atmospheric cntrl | ongoing mgmt |
| **What to save?** | artifact | idea + ancillary material & documentation |
| **Cataloging** | Individual work in hand | FRBR |
| **Later access** | Artifact & documentation | Restaging, ancillary material & documentation |

# Challenges for Digital Preservation: Standards, Architectures, & Copyright

Howard Besser, NYU Moving Image Archiving & Preservation Program

---

- http://sunsite.berkeley.edu/Longevity/
- http://besser.tsoa.nyu.edu/howard/Copyright/
- http://www.firstmonday.dk/issues/issue7_6/besser/
- http://www.cdlib.org/about/publications/
- Baca, Murtha (ed). Introduction to Metadata, Los Angeles: Getty Information Institute, 1998
  - http://www.getty.edu/gri/standard/intrometadata/
- http://sunsite.berkeley.edu/Metadata/sp2000.html
- http://www.oclc.org/digitalpreservation/presmeta_wp.pdf
- http://is.gseis.ucla.edu/us-interpares/
- http://www.niso.org/commitau.html
- http://www.ifla.org/II/metadata.htm
- METS official site: http://www.loc.gov/standards/mets
- UC Libraries Systemwide Operations and Planning Advisory Group (SOPAG) Site
  http://www.slp.ucop.edu/sopag/ for the UC Digital Preservation & Archiving Committee Final Report

# Information Commons: Public Space & the © Threat to Electronic Information Space

- http://www.info-commons.org
- http://besser.tsoa.nyu.edu/howard/Copyright/
- http://besser.tsoa.nyu.edu/howard/Copyright/commons.html

# Standards for encoding artists intentions

(group efforts w/i Cult Heritage community)

- Artists Interviews Project, Netherlands Institute for Cultural Heritage 1998-1999, **Modern Art: Who Cares** (http://www.icn.nl/english/6.4.2.html)
- TechArcheology: A Symposium on Installation Preservation (SFMOMA)
- More recent SFMOMA/Tate collaborations
- IMAP
- Guggenheim's Variable Media

# Standards, Metadata, & Best Practices-

- Risk Management
- Best Practices for Reformatting
- Identification/Provenance
- FRBR
- Technical Imaging metadata
- Discovery  & descriptive metadata
- Crosswalks
- Actors
- Other metadata (as time permits)

# Crosswalks

- mapping btwn differing metadata structures
- eliminate the need for monolithic, universally adopted standards
- focus on flexibility and interoperatiblity
- RDF-based metadata registries

# Crosswalk Example

| CDWA | Object ID | CIMI Schema | FDA | VRA Core Categories | USMARC | DUBLIN CORE |
|---|---|---|---|---|---|---|
| OBJECT /WORK (core) | Ê | Ê | Document Classification -Catalog Level (core)<br>Document Classification -Group Type | Ê | Ê | Ê |
| Object /Work-Type (core) | Type of Object | objectNAME | Document Classification - Document Type (core)<br>Purpose-Purpose (Broad) (core)<br>Purpose-Purpose (Narrow) | W1. Work Type | 655 Genre-Form | Type |
| Object /Work-Components | Ê | quantity | Document Classification -Extent | Ê | 300a Physical Description-Extent | Ê |
| ORIENTATION/ ARRANGEMENT | Ê | Ê | Ê | Ê | Ê | Description |
| TITLES OR NAMES (core) | Title | objectTitle bibliographic Title | Group/Item Identification-Repository Title<br>Group/Item Identification-Descriptive Title (core)<br>Group/Item Identification-Inscribed Title | W2. Title | 24XaTitle and Title-Related Information | TitleÊ |

# Other Elements-

- Actors Metadata
- Other Metadata
- Preserving Electronic Art

http://www.delos-nsf.actorswg.cdlib.org/

## DELOS/NSF Working Group

# Reference Models for Digital Libraries: Actors and Roles

# Other Metadata

- Description of depiction/surrogate (What VRA calls its "Surrogate Categories")
- Description of original object

- Rights and Reproduction Information
- Location Information
- VRA Core, LCSH, TGM, AAT, ULAN, TGN, DOI, <indecs>, ...

# Data Structures: The VRA Core

- 28 elements specifically for visual resource collections
- Work Description Categories-
- Visual Document Description Categories-
- http://www.oberlin.edu/~art/vra/dsc.html

# VRA Core:
# Work Description Categories

- ✿ Work type
- ✿ Title
- ✿ Measurements
- ✿ Material
- ✿ Technique
- ✿ Creator
- ✿ Role
- ✿ Date
- ✿ Repository name
- ✿ Repository place

- Repository number
- Current site
- Original site
- Style/period/group/movement
- Nationality/culture
- Subject
- Related work
- Relationship type
- Notes

# VRA Core: Visual Document Description Categories

- Visual document type
- Visual document format
- Visual document measurements
- Visual document date
- Visual document owner
- Visual document owner number
- Visual document view description
- Visual document subject
- Visual document source

# Data Value Metadata
(vocabularies)

- LCSH
- TGM
- AAT
- ULAN
- TGN
- VRA Core

# Thesaurus of Geographic Names

- over 1 million records
- hierarchical and global
- throughout history
- most records include coordinates and descriptive notes

# Identification/Provenance (Images)-

- The number of variant forms of a work can be enormous
- Image Families
- A digital image frequently has many layers of parentage
- Information about the parentage that can indicate the quality and veracity of the image (Dublin Core "Source" and "Relation")
- how to deal with different versions derived from the same scan or different encoding schemes
- Vocabulary Standards to express this

# But our ***Information Commons*** is rapidly eroding

- ***Fair use*** and ***first sale*** severely limited
- Perpetual © and elimination of ***public domain***
- Rightsholders controlling downstream use
  - To freely use it, people need to know they're not being tracked
- Licensing replacing sales
- Pay-per-use models
- Commodification of information

Even more acute with moving image works, sound, multimedia

# Importance of Public Domain and Fair Use

- What are they?
- A robust ***public domain*** is essential for a common heritage and for the creation of new works
- ***Fair Use*** is a powerful tool for both education and social commentary
- ***First Sale*** is important for social aims

Important mechanisms to assure an Information Commons of available content

# Licensing replacing sales

- We won’t own the work
- Others can control how we use the work
- The work can be withdrawn if the licensor doesn’t like how it’s used
- Licensors will try to track who views the work
- Variant forms of the work will be missing important pieces
- How will any of these works be preserved?

# Pay-per-use Models

- Discourage browsing
- Encourage viewers to follow well-established authorities
- Discourages new and innovative exploration
- Promotes best-sellers

# Commodification of Information:
## from Public Good to Tightly Controlled Commodity

- Info is a public resource, not a commodity, and like other public spaces, it is part of a "Commons"
- Librarians are traditional guardians of this public resource
- This public resource is under attack

# The erosion of the public domain-

- What is it?
- What threatens it?
- Why should we care?

# What's part of Public Domain?

**Still is**

- Air
- Sunlight
- Numbers
- God
- Ideas & Facts**

**Was**

- Water
- Land

# Public Domain
# -- a simpler explanation

- resources freely available for all members of society to do whatever they want with them
- no permissions or fees required
- no tracking of what you read or use

# Content in Public Domain

- Shakespeare
- Ballads
- Fables
- ...

# What threatens it?-

- An aggressive Content Industry
- Term extension
- Returning out-of-copyright  works back to copyright
- Mickey Mouse
- Elimination of ***Fair Use*** and ***First Sale***
- Licensing
- Other forms of Contract Laws

# Content Industry trying to control all Downstream Use

users will no longer “buy” content; they’ll “license” it

- Pushing new legislation
- Lawsuits
- Copy protection
- Contract Law
- New business models

# Hollings Bill

- All consumer devices sold in US must enforce copy-protect schemes
- What happens to Fair Use rights?

# www.digitalconsumer.org advocates confirming consumer rights to:

- Time-shift content
- Space-shift content
- make back-up copies of content
- Translate content into diff formats (e-book becomes audible for blind)
- No technological barrier should deprive one of their Fair Use rights

# Content Industry Plans

- [advocating legislation that] “guarantees publishers’ control of not only the integrity of an original work, but of the extent and duration of users’ access to that work, the availability of data about the work and restrictions on forwarding the work to others” -- Peter Chernin, News Corp President (owner of Harper-Collins) quoted in Publishers Weekly, May 2001

# Strong Content Control Threatens Creativity

**Imagine Creators:**

- Having no public domain to draw upon
- Having to negotiate rights for every clip, every drawing, every still image, every sound sample
- Having to renegotiate all these rights every time they redistribute it in a different form or in a different media

# Strong © protection/DRM can inhibit Teaching

- Sound or film clips
- dist learn violating ©-
- Teach Act-
- My stories-
- Threats for posting readings
- Curricular Materials-

# TEACH Act (2003)

## Technology, Education, And Copyright Harmonization

- Fixes © exemptions dissolved by 1998 DMCA
- Allows classroom/distance use of © material provided certain conditions are met:
  - Under actual supervision of instructor as part of an accredited nonprofit educational institution
  - As an integral part of a class session
  - Access controls protect against non-student viewers
- Only certain types of Works
  - Exemption doesn't cover works marketed for mediated classroom
  - Can convert analog to digital only if no digital is available (or digital version is protected)
  - More open about using nondramatic works than using dramatic works

# Other serious erosions

- UCITA--shrink-wrap
- Database Treaty
- Anti-Circumvention
  - Felten & SDMI challenge
  - Sklyarov-

# Criminalizing © violations
# Dmitri Sklyarov jailed

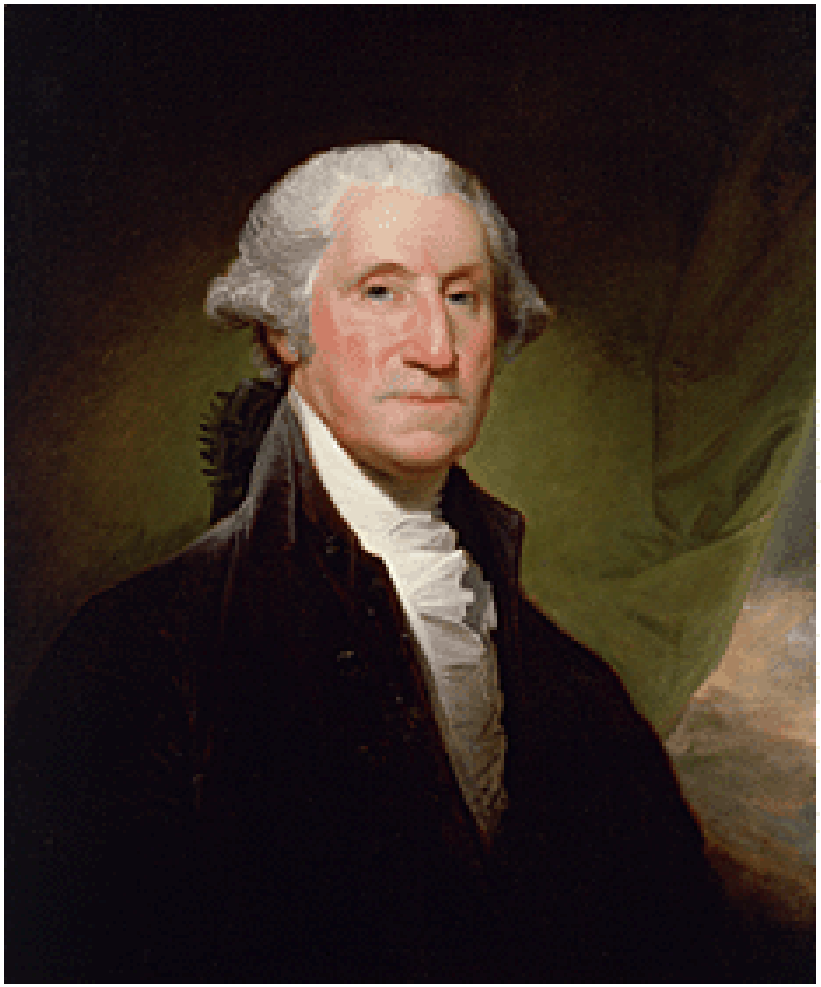
# Database Protection Legislation

- The educational exemptions in this bill “appear too narrow to support current university information-use practices”, and it appears that ANY claim of “market harm” could nullify the fair use exemptions.
- --Dr. Debra W. Stewart, testifying Feb 16, 1998 on behalf of the Association of American Universities in front of the House Judiciary Subcommittee on Courts and Intellectual Property considering proposed Database Protection legislation

# Content Industry Wants

- A world of consumers
- Minimal interference from content producers who may raise alternative voices
- Culture becomes a mere commodity, and discourse around it is tightly controlled
- A world where they maintain power, control, ...

# Wa$hington



*Vaughn Portrait*
Gilbert Stuart, 1795