Data Preservation: What’s New?

Victor McCrory
Convergent Information Systems Division
National Institute of Standards and Technology

Problem: The Digital Content Dilemma

Lack of interoperable standards and guidelines for content types, transfer, and storage lead to increased cost for content creation and slow the development of future e-commerce applications in learning, healthcare, and government.
The Digital Dilemma: Issues Facing the Industry for Digital Content

- Guidelines and standards for archiving data & storage media
- Copyright ownership and management
- Standards for content exchange
- Technologies and architectures to provide low-cost anonymity and security
- Fidelity of content manifestation

### Digital Image File Comparison

<table>
<thead>
<tr>
<th>File Type</th>
<th>Support</th>
<th>Uses</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Interoperability</th>
<th>bitrate</th>
<th>Standards</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIF (Graphics Interchange Format)</td>
<td>No browser plug-in is required, but since the GIF is widely supported in most browsers, this does not pose a problem.</td>
<td>Can be used for animation, icons, and simple graphics.</td>
<td>Highly compressed, supports multiple image formats.</td>
<td>Limited support for animation.</td>
<td>GIF is an older format.</td>
<td>GIF is an older format.</td>
<td>No standard for animation.</td>
<td><a href="http://www.w3.org/Photo/HTML/Graphics/GIF.html">http://www.w3.org/Photo/HTML/Graphics/GIF.html</a></td>
</tr>
<tr>
<td>PNG (Portable Network Graphics)</td>
<td>Uses the advanced compression capabilities of PNG, for example, supports multiple image formats.</td>
<td>Useful for lossless compression and transparency.</td>
<td>Better compression than GIF, supports alpha channel, and transparency.</td>
<td>Not widely supported, though the W3C is working on it.</td>
<td>Works well with other image formats due to its lossless compression.</td>
<td>PNG is an open standard.</td>
<td>PNG is an open standard.</td>
<td><a href="http://www.w3.org/Graphics/PNG/">http://www.w3.org/Graphics/PNG/</a></td>
</tr>
<tr>
<td>SVG (Scalable Vector Graphics)</td>
<td>Uses vector data, which can be scaled without losing quality.</td>
<td>Useful for creating scalable graphics.</td>
<td>Very scalable, supports transparency and interactivity.</td>
<td>Fewer programs support SVG.</td>
<td>Supported by Adobe Illustrator, CorelDRAW, Inkscape, and many other programs.</td>
<td>SVG is an open standard.</td>
<td>SVG is an open standard.</td>
<td><a href="http://www.w3.org/Graphics/SVG/">http://www.w3.org/Graphics/SVG/</a></td>
</tr>
<tr>
<td>TIFF (Tagged Image File Format)</td>
<td>Supported by most applications, can be embedded into a document.</td>
<td>Ideal for applications that require high-quality images.</td>
<td>Support for many image formats, including RGB and CMYK.</td>
<td>Large file sizes.</td>
<td>Very flexible with multiple formats.</td>
<td>1 to 20 bits of information per pixel.</td>
<td>TIFF is an open standard.</td>
<td><a href="http://www.w3.org/Graphics/SVG/">http://www.w3.org/Graphics/SVG/</a></td>
</tr>
</tbody>
</table>
Legislative & Government Initiatives:

♦ Health Insurance Portability and Accountability Act (HIPAA) – compliance by April 14, 2003
♦ National Digital Information Infrastructure and Preservation Program (NDIIPP) – Library of Congress
♦ ISO-TC 171/SC 1N 20 – Care & Handling of Optical Media

NIST Digital Data Preservation Laboratory

♦ Partner with industry to work towards voluntary industry standards conformance, and convening the industry to develop sector solutions.
♦ Development of standard testing protocols for reliability of optical disc media.
♦ Development of standard testing protocols and software for drive interoperability for writable optical disc formats.
♦ Research new media candidates for increased optical storage density.
♦ Evaluate new technology trends for applications at NIST and other Federal agencies.

Provide standards, test methodologies, and test beds to ensure content is available now and in the future.
NIST Digital Data Preservation Laboratory: Customers and Applications

- Content creators and providers.
- Optical disc & drive manufacturers.
- Archival storage & data preservation.
- High density storage and applications.
- Digital cinema & interactive DTV.
- Digital libraries & electronic books.
- Digital Rights Management.
- End-users.
**NIST Optical Disc Compliance Test Program**

**Problem Statement:**
- No mechanism available to the consumer to test their drive for MultiRead compliance

**NIST Response:**
- NIST developed its own test, based on the MultiRead specifications, for customers to test drives for compliance. The test comprises of the following:
  1.) generation of test pattern,
  2.) burn the pattern test disc,
  3.) test drive for MultiRead compliance using test disc and show and print the results of the test.

---

**The NIST Optical Disc Compliance Test:**

1.) Create source (pattern) files.
2.) Burn the source files to the test disc
3.) Test for MultiRead compliance using the test disc created (based on MultiRead Test plan).
4.) Prints results outlining where (if any) failures occur.
5.) Help and Information.
Standards and Interoperability

Test Functions

- Can drive...
- Read media of different reflectivity?
- Read all User Data Blocks on Data Tracks?
- Read different write methods?
- Read tracks using different Addressing methods?
- Support Multimedia Command set (MMC)?
- Access Reading Link Blocks?
- Detect and identify errors?

Standards and Interoperability

Current Interoperability Status

<table>
<thead>
<tr>
<th>Media</th>
<th>CD-ROM</th>
<th>CB-R</th>
<th>CD-RW</th>
<th>DVD-ROM</th>
<th>DVD-R</th>
<th>DVD-RAM</th>
<th>DVD+RW</th>
<th>DVD+RW</th>
<th>PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-ROM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CD-R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CD-RW</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DVD-ROM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DVD-R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DVD-RAM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DVD+RW</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DVD+RW</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

* = YES  R = RECOMMENDED  O = OPTIONAL

ACHIEVED BY DRIVES WITH "MULTIREAD" CAPABILITY

OSTA (www.osta.org)
<table>
<thead>
<tr>
<th>Organization</th>
<th>Standard/Collaboration</th>
<th>NIST Role/Deliverable</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical Storage Technology Association (OSTA).</td>
<td>MultiRead: defines the parameters necessary for optical devices to read discs created in CD formats.</td>
<td>Develop test to ensure MultiRead compliance.</td>
<td>MultiRead test (CISD website), MultiRead 2 (DVD) test under development.</td>
</tr>
<tr>
<td>Joint Technical Commission: Care and Handling of media.</td>
<td>ISO TC 171/SC 1N 120 &quot;Verification of information on CD's&quot;:</td>
<td>Member of Commission. Provided help and feedback on proposal.</td>
<td>ISO working draft.</td>
</tr>
<tr>
<td>Warner Advanced Media Operations (WAMO).</td>
<td>Reflectance Measurement for DVD.</td>
<td>CISD measured the reflectance to an accuracy of 1%.</td>
<td>Result used as the standard reflectance for all DVD discs.</td>
</tr>
<tr>
<td>Library of Congress (LOC) / National Archives</td>
<td>Promotion of importance of Data Preservation for Archival purposes.</td>
<td>Proposed establishment of Data Preservation Lab at NIST.</td>
<td>In proposal phase.</td>
</tr>
</tbody>
</table>

**Preservation of Digital Data**

*Preservation of Digital Information*

- Digital discs have large capacity, fast access rate, longer life expectancy and lossless information transfer.

- Lifetime of media is limited due to deterioration of material.

- Longevity of media also determined by the obsolesce of playback hardware and software.

- Compatibility and interchangeability of content on different disc types.

- This effort is very important for long-term preservation of digital information for archival purposes as well as for National Security.
Preservation of Digital Data

Example of Result:

Defect Detection and Analysis
With some hardware improvement this method can be used to accurately characterize DVD discs.

Preservation of Digital Data

NIST effort

Independent measurement of life expectancy:
- Chemical deterioration/aging.
- Physical damage.
- Content loss in operation.
Rating of media (EOL).
Development of universal software for reading various digital formats.
Promote the development of standards for digital storage in digital discs.
DVD Reliability/Compatibility study.
MultiRead test.

Environmental Chamber for Reliability of DVD Media
Optical Disc Reliability Testing

- Stimulate the use of Optical Discs for Preservation:
  Much of the original digital information produced today is placed on CD-R and DVD-R media (valuable experimental data from a scientist, for example) and therefore that is where we concentrate our efforts.
- Identify the discs that are suitable for archiving:
  This recommendation can be used by archivists and librarians as a disc selection guide.
- Develop a methodology to test existing discs:
  Allow archivists and librarians (and others) to test the valuable discs already in their collection.

Tools Used For Testing

- Temperature/Humidity and Light Chambers
  - CISD developed light chamber is one of the only of its kind in the world.
- Optical Disc Microscope Testbed
- CATS SA3 Compact Disc Analyzer
- CD Associates DVD Analyzer
**NIST/CISD Light Chamber**

**Optical Disc Test Equipment**

- A3 Compact Disc Analyzer
  - Provides CD parameters after
  - temperature and/or humidity effects

- CD Associates DVD Analyzer
  - Provides DVD parameters after
  - temperature and/or humidity effects
Different Manufacturer’s Media Response to Temperature & Humidity

Temperature Effects on Optical Discs
Humidity Effects on Optical Discs

![Humidity Effect Graph]

HDSA/NIST
Data Preservation Lab
**HDSA/NIST**

**Data Preservation Lab**

- **Goal:** Create a real life demonstration facility for present and future end users of these systems.

- **Goal:** Highlight the capabilities and limitations of different jukebox types and systems.

- **Goal:** Improve jukebox and library interoperability, performance and transfer rates.

- **Goal:** Study interchangeability between storage systems, media types, and storage management.

- **Goal:** Work with industry (HDSA) to create a Storage Technology Tool that evaluates different storage media.

---

**Overview of HDSA/NIST Lab**

*Layout of the NIST/HDSA High Density Storage Laboratory*
Research and Services

Future Projects

- Evaluation of new technology such as FMD:
  - Fluorescent Multi-Layered Disc-uses dye polymer as storage medium
  - Storage system player & media compatibility
  - Expected lifetime measurement when exposed to sun/ambient room light.
- Standardization of Mass Storage:
  - Promote mass storage systems/sub-systems interoperability
  - Develop testing protocols for performance of mass storage systems and data exchange

Advanced Coding for Data Preservation

- Problem Statement:
  - To evaluate the performance of error correction codes in the recovery of information lost on optical media due to aging or some catastrophic event.
Advanced Coding for Data Preservation

- Identify the parameters related to data error on optical discs.
- Measure these parameters before and after accelerated aging.
- Process the data by both conventional ECCs and more advanced ECCs.
- Develop standardized formats for long term preservation of valuable data.

Medical Image Quality
Analog Preservation: Preserving the Constitution

The second page of the Constitution is securely displayed in a new, titanium-framed encasement made at NIST

Solution: CISO Provides Interoperable Tools and Guidelines for Content Management (Standards, Measurements, Interoperability needs)

<table>
<thead>
<tr>
<th>Content Creation AND Packaging</th>
<th>Content Conversion AND Interoperability</th>
<th>Content Transfer AND Delivery</th>
<th>Content Manifestation AND Consumption</th>
<th>Content Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMI Needs</td>
<td>SMI Needs</td>
<td>SMI Needs</td>
<td>SMI Needs</td>
<td>SMI Needs</td>
</tr>
<tr>
<td>1A. Inter-Content Interoperability (jpeg,mp3,dct)</td>
<td>1A. Interoperability of DRM Solutions</td>
<td>2A. Interoperability of DRM Solutions</td>
<td>3A. Interoperability of DRM Solutions</td>
<td>4A. Interoperability of DRM Solutions</td>
</tr>
<tr>
<td>1B. Intra-Content Interoperability (jpeg vs. tif)</td>
<td>2B. Performance and Usability for Various DRM Solutions (e.g. Biometrics, etc.)</td>
<td>3B. Content Integration during transmission</td>
<td>4B. Quality standards for tools on various platforms</td>
<td>5B. Interoperability for writable media</td>
</tr>
<tr>
<td>Architecture guidelines for various content types</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Red: Short-term, immediate benefit
Yellow: Medium-term benefit
Green: Long-term benefit

End-User Applications
- E-Learning
- Healthcare
- E-Government
- E-Audio Maintenance
- E-Commerce

High Usability and Wide Deployment

Lower Costs
Higher Interoperability
NIST Industry Collaboration in Data Preservation

Conclusion

- NIST has an extensive program involvement in digital data preservation:
  - media reliability & interoperability
  - storage system performance metrics
  - member of National Digital Strategy Advisory Board for NDIIPP
- For more information:
  victor.mccrary@nist.gov
  Phone: 301.975.4321
  www.itl.nist.gov/div895