Write Once, Read Forever (WORF): Low-energy storage in perpetuity of high-density, multi-state data

Richard J. Solomon, University of Pennsylvania, Philadelphia, PA, USA and Creative Technology LLC, Monson, MA, USA; Melitte Buchman, New York University, New York, NY, USA; Clark Johnson, Creative Technology LLC, Madison, WI, USA; Eric Rosenthal, Creative Technology LLC, Morganville, NJ, USA, and New York University, New York, NY, USA; Jonathan M. Smith, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Successful archiving of data requires a storage medium capable of persisting for times measured in centuries, and providing an absolute trust in supportable and permanent hardware. While all media decay, data on magnetic media, solid-state drives and conventional optical disks must be cyclically refreshed over relatively short time frames, requiring energy and labor resources. We propose a "Write-Once, Read Forever" (WORF) module specifically engineered for long-term preservation of digital data. It uses a novel high-density data storage medium based on silver halide, which has been demonstrated to last for more than a century under normal ambient environmental conditions. Once data is written to WORF, energy is needed only for reading—no periodic refresh is necessary, and data is both immutable and truly permanent. Human readable text and images are embedded in the WORF module adjacent to the digital data. This text and imagery contains meta-information about the media's content, and instructions for decoding for future generations. WORF digital data is stored as microscopic, metallic interference gratings (representing wavelengths or "colors") embedded in a modern, super-resolution, dye-free, photosensitive emulsion. Wavelengths encode multiple states per data region; current spectroscopic technology makes 400 states per 2 micron diameter data region feasible. Multi-state data architecture within each domain enhances data integrity, error-checking, and accelerates writing and reading for the entire media module.

The problem

Current magnetic and optical digital media are not ideal for archival data storage, due both to excessive energy requirements and inherent environmental deterioration [1, 2]. Over the long term, all storage media deteriorate. Digital data, as copies must be perfect, is particularly vulnerable to decay, requiring strict environmental control of temperature and humidity, consuming continual energy resources. Periodic regeneration or refreshing of data may maintain a perfect copy due to its digital state, albeit at significant overhead over the time frames required for archival work.

Lifetimes without regeneration of conventional media range from as little as a few months to at most a few decades (see Fig. 1 for comparisons of current media to some exemplary archival materials). Magnetic media are subject to thermal decay and magnetic field data erasure, exacerbated by environmental conditions in large data storage facilities [3, 4, 5]. Most magnetic data storage requires constant monitoring of critical environments, continual and sometimes heavy energy consumption, and periodic regeneration for archival purposes [6]. Solid-state (SSD) memory is more vulnerable than a magnetic device, and like magnetics, require periodic refreshes—with some manufacturers' recommendations as short as three months—with long-term serious implications for energy consumed [7, 8, 9, 10]. Legacy optical disks (CD/DVD/Blu-ray) store data either as microscopic holes etched into an aluminum film, or via a photosensitive dye susceptible to relatively short-term degradation [11, 12, 13]. Over time, aluminum film oxidizes, destroying the recorded data despite a protective layer covering the disk. No protective layer is totally impervious to penetration by the atmosphere and pollutants, so eventually the data will be lost. Dyes inherently fade over time, and those used on these disks are sensitive to strong light, especially sunlight [14]. Moreover, besides oxidation, the metal films used are subject to attack from microorganisms including fungi and bacteria [15].

The bottom line for archival data storage is that vast amounts of important scientific and societal information may be lost due to fragile media storage technologies: we may become a generation making decisions on "big data," all of which will be lost!

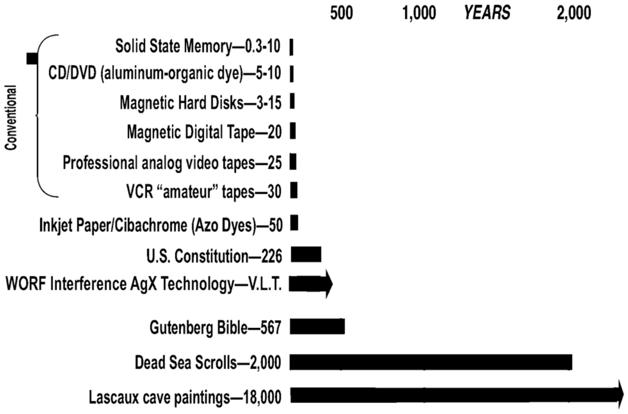

***Figure 1**. Estimated lifetimes of Selected Media in Ambient Room Temperature (unpowered and unrefreshed). WORF should last as long as the U.S. Constitution based on lifetimes of AgX extant media, and likely longer.*

A Permanent Archival Solution

Unlike conventional contemporary media, our "Write-Once, Read Forever" (WORF) module has been designed specifically to persist for a minimum of several centuries based on known properties of silver halide chemistry. This is a solution intended not for lowest cost per bit or enduring artificial lifecycle tests, but for true permanence. Further, since instructions for reading WORF data are included on the media, in human readable form, this is an instance of a reading device that can be realized even in the far future should the original devices be lost: the information for creating a device and decoding the formats—a digital "Rosetta Stone"—is embedded with the data.

WORF modules in the archival chain require no power for refreshing or regeneration, and only ambient room temperatures and humidity for storage. The data stored on a WORF module cannot be altered—it is truly permanent; no network authentication for checksums is necessary to confirm the accuracy of multiple copies, an essential requirement for distributed archival data.

Silver halide

WORF stores digital data on a special type of extremely high resolution, photosensitive, silver-halide (AgX) suspension, known as a Lippmann emulsion. Current digital storage appliances, with their volatility, energy consumption, and requirements for duplication, etc., are in stark contrast with the stability of properly processed, photosensitive, silver-halide media [16]. The preservation community used silver halide for microfilm, one of the first international standards for long-term preservation [17, 18]. AgX life expectancy is also evident from photographs over 150 years old that show no deterioration. Finally, the silver ion (Ag^+) is hostile to most bacteria and other microorganisms that attack conventional media [19, 20].

WORF uses modernized Lippmann emulsions, which were originally used by him and others a century ago to store full-spectrum color photographs via microscopic interference gratings that physically capture the standing waves of all of the colors in a scene [21]. Full-color Lippmann plates using early panchromatic chemistry are extant from the 1890s showing no deterioration of color or image [22]. More sensitive Lippmann emulsions have subsequently been formulated with modern chemistry for electron microscopy and holography [23, 24, 25]. The ultimate density of such modern silver halide emulsions has yet to be exploited for storing data. WORF reworks the system to store extremely narrow-band frequencies in the form of interference patterns encoding multi-state digital data.

For conventional photographic applications, microscopic silver halide grains (in particular, silver bromide, AgB, or silver iodide, AgI) are embedded in suspensions (emulsions on a substrate such as glass or transparent film) to capture varying light intensities creating a monochromatic image. Grain size and emulsion thickness vary depending upon sensitivity (film "speed") and other factors relevant to photographs [26]. Modern color films apply a set of tri-chromatic dye couplers embedded in the emulsion to re-create a practical subset of the visible spectrum; these dyes eventually fade in ambient temperatures [27], and are therefore unsuited for archival systems.

Lippmann's technique for full-spectrum color photography was quite different from color film and digital processes used today. Lippmann emulsions capture visible frequencies directly, storing the actual wavelengths of different colors by exposing archival (monochromatic) silver halide grains in a unique arrangement to construct a multitude of microscopic diffraction gratings vertically positioned in the emulsion [28, 29]. No unstable dyes or pigments are involved. The gratings are created by standing waves reflected from a substrate creating interference patterns representing the precise wavelengths of the colors in the scene. White light, illuminated at the correct Bragg angle for the gratings, through the exposed and developed emulsion, re-creates the precise wavetrains or frequencies originally exposed. Figures 2 and 3 illustrate standing waves for two frequencies overlaid on an actual emulsion microphotograph cross-section [30].

To create these gratings, Lippmann silver halide grains must be smaller than the shortest wavelength of visible light (*i.e.*, the wavelength of blue ≈400 nm), and to capture a useful vertical grating, the depth of the emulsion layer must be at least twice the longest wavelength (*i.e.,* red ≈800 nm), or about 2 microns thick on its substrate [31]. Most conventional photographic film emulsions are thinner than this, and for sensitivity reasons have larger silver halide grains, so thus are not suitable for storing interference standing waves.

Why Now?

Lippmann emulsions have been explored and refined for over a century, but with WORF, three rapidly advancing technologies have converged, offering a unique solution for extremely long-term digital archiving:

1. modern wave-based writing and sensing devices;
2. advanced mechanisms and photosensitive media for storing spectrographic interference gratings; and,
3. implementation of multiple-state data storage architectures.

By combining these techniques, WORF resolves the long-standing dilemma of how to conserve digital archives for future generations, while minimizing labor and energy.

How WORF works

Lippmann used extremely slow panchromatic emulsions of the day, and plates dipped in mercury for the interference reflecting mirror. WORF uses more sensitive contemporary emulsion chemistry specifically crafted for higher-intensity data writing and reading mechanisms, and more modern coatings for reflectivity both for writing and reading, such as a metallic deposition. WORF repurposes the Lippmann emulsions, and its concomitant interference process, applying contemporary electro-mechanical, optical, and spectroscopic components for writing and reading high-density digital data.

The necessary precision for WORF is well suited to off-the-shelf mechanisms currently used for magnetic or optical disk drives. These mechanisms are enhanced with spectrographic write and read heads using previously developed technology for writing and reading narrow spectra (see discussion and references for Figs. 2 & 3). Within a domain approximately 2 microns in diameter, current spectroscopic technology permits up to 400 frequency states (within the visible range, that equates to 400 different "colors" segmented 1 nm apart). Further refinements may approach 4000 states per domain by implementing optical frequency/amplitude constellations similar in concept to tuning in data modems.

Interference patterns are created by standing waves reflected from the suspension substrate coated with a reflective deposition, or via a mis-matched index of refraction (Fig. 2). The multi-frequency light source can be a deformable micro mirror device (DMD) configured to generate any 1 nm wide frequency within 1/30,000 sec. [32]. Reading the interference frequencies representing data is the reverse: white light transmitted at the critical Bragg angle for the grating is diffracted into the various waveforms expressing the multi-state data (Fig. 3) directly into a full-spectrum detecting apparatus [33]. Rapid random access data retrieval may be performed by detecting all the frequencies in an array from the substrate simultaneously by using a computer-controlled direct-frequency light detector [34, 35, 36].

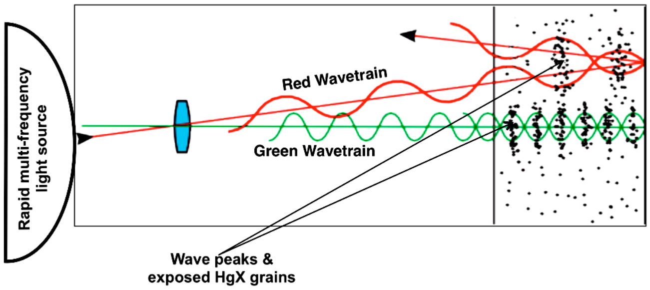

Figure 2. *Writing WORF Data. Illustration superimposed on a microphotograph [30] of a Lippmann emulsion showing AgX grains, corresponding to peaks in colored wave trains, that are exposed creating spectroscopic gratings embedded in the emulsion as the light reflects from the suspension base interfering with the forward waves; the gratings are essentially standing waves precisely representing the frequencies stored. Longer wavelengths (e.g., red) expose a coarser grating; shorter wavelengths (e.g., green) expose a finer grating. Different colors emanating from the rapidly modulated multi-frequency light source [e.g., Refs. 32, 36] allow multi-state data to be stored per data region.*

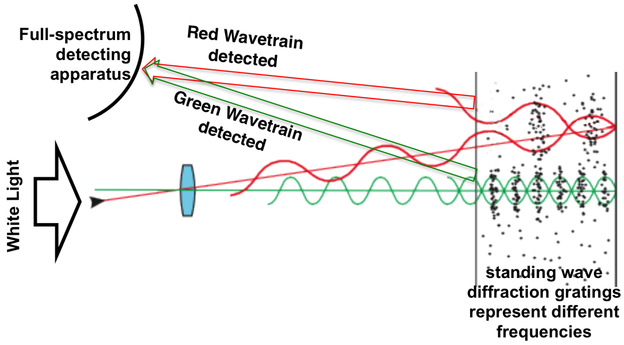

Figure 3. *Reading WORF Data. Illustration superimposed on a microphotograph [30] cross-section of a Lippmann emulsion showing standing waves embedded as AgX diffraction gratings in the emulsion. The gratings modulate a white light source so that the resulting reflected complex multiplexed wavetrains, read by the full-spectrum detecting apparatus [e.g. Refs. 33, 34, 36], extract the stored multi-state data. The detecting apparatus can be an array of narrow-band frequency detectors [e.g., Refs. 34, 35] capturing stored data simultaneously for rapid random access.*

Data Architecture & Multi-state Bitstrings

Data storage systems using photosensitive media (including dye-based CD/DVDs, as noted) have been proposed before, but have not been suitable for long-term data archival purposes for various reasons:

1. media that used various polymers for rapid recording degraded readily;
2. data on linear tape systems cannot be rapidly retrieved making random-access practically impossible;
3. holographic systems—which store spatial information in the amplitude domain rather than in the frequency domain—create an interference effect by adding the wavelengths from two laser beams, so a small region of a hologram can reproduce the entire dataset—however, resolution is very low and hence impractical for high-density archiving; and,
4. all these prior systems proposed storing binary data, which complicates error checking, as well as exhibiting inadequate densities for practical archiving.

In addition to increased density, WORF's multi-state data has several advantages compared to amplitude (intensity)-modulated, 2-state dots:

1. writing and retrieval is frequency-modulated (within one octave), reducing noise and consequent error-correction complexity [37];
2. multi-state data architecture at each domain enables an integral checksum to be implemented *within* the domain, instead of for the media as a whole (a major defect of holographic storage);
3. error correction is enhanced, compared to magnetic or solid-state storage, since the standing wave gratings are permanent, stable and cannot change or flip; and,
4. unlike microfilm or microfiche storing alphanumeric characters or low-resolution images, multi-state data can be directly translated by existing computer software into arbitrary digital formats.

A future task is to develop a novel computer architecture encoding these multi-state bitstrings, rather than employing legacy binary arithmetic for data storage. And, because the media is essentially optical—albeit very different from other optical data storage media, including CD/DVDs—it can readily combine a physical component for human readability and decoding in the distant future with its digital component (see Fig. 4).

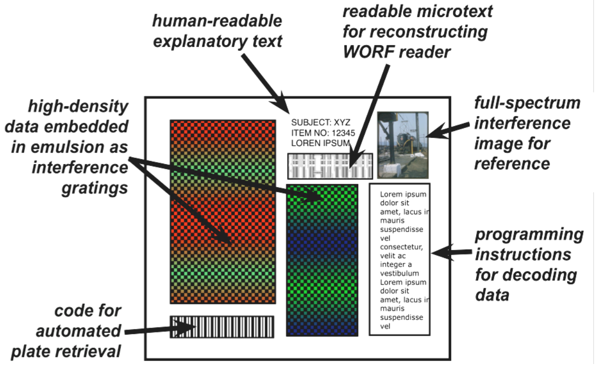

Figure 4. *One possible media format with meta-data for decoding in the distant future, and instructions for building a reading device if necessary. The required format for WORF is agnostic. Other formats could be disks, wallet cards, memory sticks, etc.*

The WORF data storage architecture has two logical layers. The lower of the two defines units of storage based on the digital encoding of data, and provides data protection via the use of Cross-interleaved Reed-Solomon Coding (CIRC) as used in compact disks [38]. CIRC coding provides desirable protections against both random errors and long burst errors such as scratches. This layer provides the illusion of an array of bytes to the upper layer. The upper layer structures the lower layer's array of bytes as a UNIX file system [39], including redundant metadata describing the mapping between named objects and the units of storage ("blocks") provided by the lower layer. As the WORF medium is immutable, the redundancy is primarily employed to ensure the integrity of copies. Advantages of using a conventional file system include high quality open source implementations and available documentation, both of which can be included as WORF metadata.

Digital Preservation for the Long Term

WORF is intended as a replacement archival module for the current volatile backup tier in the digital storage hierarchy. WORF simply replaces Tier Three in the digital preservation chain (see figure 5) [40]. WORF becomes a true archival tier, not requiring refreshing or power, and therefore conserving energy over the long term. No major reconstruction of the preservation chain is necessary. With WORF, storage infrastructure becomes simpler and can be located anywhere, instead of near abundant cooling water and power generating stations. Since WORF media cannot be changed after writing, network authentication among multiple backup sites is unnecessary for validation. As noted, WORF substrates can include metadata, indexing codes, and human-visible images; instructional text in microtext can be included on the medium to enable data decoding and the construction of readers in the far future should such apparatus and techniques be lost to time.

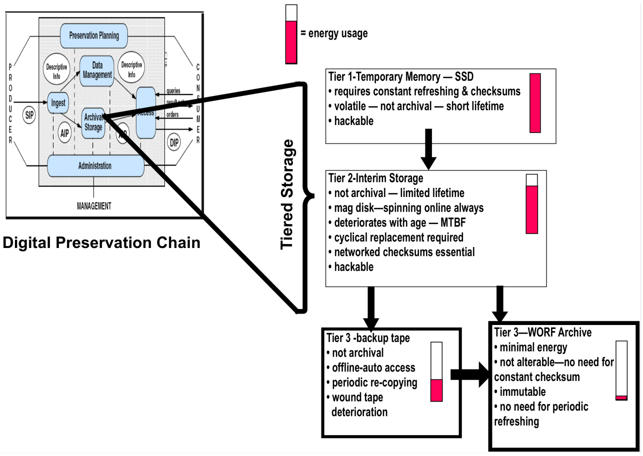

***Figure 5.** The WORF Module in the Digital Preservation Chain [40] replaces current modules — tape, CD/DVDs, magnetic hard disks — with a true archival media that does not require periodic refreshing and can be stored in ambient room temperature. Energy usage is minimal, and constant attention is not necessary, saving significant labor inputs.*

Currently the deep storage of most digital archives is LTO ("Linear Tape-Open") backup tapes. These tapes are used partially because they align with clearly stated standards that museums, libraries and other communities depend on. The Open Archival Information System (OAIS) reference model [41] is one such standard. In the OAIS model active storage function is required to allow functions such as receiving archival information packages, error checking, and access functions. However it is the OAIS archival storage function that includes assigning digital files permanent storage. This "deep" storage is also part of another important standard LOCKSS [42] ("Lots of Copies Keep Stuff Safe") whose mandate includes perpetual community access, via redundancy, affordability and sustainability as part of an archive.

Although LTO tapes are currently used for this archival storage function they are expensive, require constant attention and consume high rates of energy. Over the lifetime of a long-term archive numerous different individuals and institutions may work together to maintain archives within the vicissitudes of business cycles, personnel and institutional turnover. In the face of change, cost, and technological evolution the dark archive must remain. To have absolute trust in the archive there must be absolute trust that the archive hardware is supportable and permanent. Clearly building an OAIS compliant repository on less than trusted or supportable architecture is the issue we now all face.

References

[1] D. Rosenthal, *et al,* The Economics of Long-Term Digital Storage, Proc. The Memory of the World in the Digital Age: Digitization and Preservation, UNESCO (2012) pg. 523, and *passim.*

[2] D. Rosenthal, "Keeping Bits Safe: How Hard Can It Be?," ACM Storage, 8, 10 (2010).

[3] R. Skomski and D. Sellmyer, "Intrinsic and Extrinsic Properties of Advanced Magnetic Materials," in Yi Liu, ed., Handbook of Advanced Magnetic Materials (Springer, 2006) pp. 1-57.

[4] D. Bajalan, Nano Segregation Effects on Nano Magnetic Properties in Multi-Layer Thin Films, Prog. In Electromagnetics Research Symposium, Hangzhou, China, pp. 674-675. (2005).

[5] D. Weller and T. McDaniel, "Media For Extremely High Density Recording," in D. Sellmyer, Advanced Magnetic Nanostructures (Springer, 2006) pp. 304-307, and *passim.*

[6] D.Thompson and J. Best, "The future of magnetic data storage technology," IBM J. Res. & Dev., 44, 3, pg. 311. (2000).

[7] M. Cornwell, "Anatomy of a Solid-state Drive," ACM Networks, 10, 10 (2012).

[8] D. Lorenz et. al, "Efficiently analyzing the impact of aging effects on large integrated circuits," Microelectronics Reliability, 52 (2012) pp. 1546–1552.

[9] Fujitsu PRIMERGY Servers Solid State Drives – FAQ, <http://ts.fujitsu.com/primergy> (2013) pg. 15.

[10] S. Kumar and R. Vijayaraghavan, Dell Solid State Drive (SSD) FAQ, <www.dell.com/downloads/global/.../en/solid-state-drive-faq-us.pdf> (2011) pg. 6 & *passim.*

[11] J. Zheng, NIST/Library of Congress (LoC) Optical Disc Longevity Testing Procedure, NIST Spec. Pub. 500-263 (2005).

[12] O.Slattery et. al, "Stability Comparison of Recordable Optical Discs—A Study of Error Rates in Harsh Conditions," J. Res. NIST, 109, 5 (2004).

[13] F.Byers, "Care and Handling of CDs and DVDs — A Guide for Librarians and Archivists," NIST Spec. Pub. 500-252 (2003).

[14] J. Iraci, "The relative stabilities of optical disc formats," Restaurator, 26, 2 (2005) pp. 140-141.

[15] X. Bosch, "Fungus eats CD—Spores bore holes in compact disks, rendering them useless," Nature, doi:10.1038/news010628-11 (2001).

[16] H. Wilhelm et. al, Long-Term Preservation of Photographic Originals and Digital Image Files in the Corbis/Sygma Collection in France, Proc. IS&T 2008 Archiving Conf., Tables 1 &2 (2008).

[17] American National Standards Institute, Recommended Practice for Microfilming Public Records on Silver-Halide Film, ANSI/AIIM Standard No. MS48-1999 (1999).

[18] H. Wilhelm and C. Brower, The Permanence And Care Of Color Photographs: Traditional And Digital Color Prints, Color Negatives, Slides, And Motion Pictures (Preservation Pub. Co., Grinnell, Iowa, 1993) pp. 92-93.
[19] C. Chambers et. al, “Bactericidal Effect of Low Concentrations of Silver,” J. Am. Water Works Assoc., 54, 2 (1962) pp. 208-216.
[20] R. Belly and G. Kidd, “Silver Resistance In Microorganisms,” Dev. Ind. Microbiol. 23 (1982) pg. 567 fn.
[21] G. Lippmann, “La Photographie des Couleurs,” Compte Rendes a l’Academie des Sciences, 112 (1891) pp. 247-275.
[22] W. Alschuler, On the physical and visual state of 100-year-old Lippmann color photographs, Proc. SPIE 3358 (1997) pp. 84-94.
[23] H. Bjelkhagen, Silver-Halide Recording Materials for Holography and Their Processing, (Springer, 1995) ch. 2.2.2, pp. 37*ff.*
[24] C. Rich, et. al., Lippmann photographic process put to practice with available materials, Proc. SPIE 2688 (1996) pp. 88-95.
[25] H. Philippaerts, et. al., Photographic Lippmann Emulsions, U.S. Patent No. 4,202,695 (1980).
[26] T. James, The Theory of the Photographic Process (4th Ed., Macmillan, 1977) ch. 3, sect. D.
[27] H. Wilhelm, and C. Brower, *op. cit.* (1993) ch. 2 & *passim*.
[28] L. Mannheim, ed., Photography Theory And Practice (Amphoto, 1970), vol. I, sect. 58.
[29] R. Wood, Physical Optics (Macmillan, 1934) pp. 214-217, figs. 137-138.
[30] Microphotographs from < http://sechtl-vosecek.ucw.cz/en/expozice5.html#color>.
[31] L. Mannheim, *.op.cit*.
[32] E. Rosenthal, R. Solomon, C. Johnson, “Full Spectrum Color Projector,” U.S. Patent No. 6,985,294 (2006).
[33] E. Rosenthal, R. Solomon, C. Johnson, *“*Full Spectrum Color Detecting Pixel Camera,” U.S. patent No. 7,437,000 (2008).
[34] E. Rosenthal, R. Solomon, C. Johnson, “Electromagnetic Spectral-Based Imaging Devices And Methods,” U.S. Patent No. 7,521,680 (2009).
[35] E. Rosenthal, R. Solomon, J. Smith, C. Johnson, “Software-defined Lensing,” U.S. Patent Nos. 8,507,836 (2013).
[36] E. Rosenthal, R. Solomon, C. Johnson, “Wave-Based Optical Data Storage,” U.S. Patent Office Provisional Application: No. 61/519,980 (2011).
[37] Christiansen, D. ed., et.al., Electronic Engineers’ Handbook (McGraw-Hill, 4th Ed., 1996) sects. 4.10 & 4.11.
[38] K. Immink, “Reed-Solomon Codes and the Compact Disc,” in S. Wicker and V. Bhargava, Eds., Reed-Solomon Codes and Their Applications (IEEE Press, 1994).
[39] M. McKusick, et. al, “Fast File System for UNIX,” ACM Trans. on Computer Systems 2, 3 (1984) pp. 181-197.
[40] “Digital Records Pathways: Topics in Digital Preservation,” Draft (InterPARES/ICA, July 2012) p.20.
[41] <http://oclc.org/research/publications/library/2000/lavoie-oais.html>
[42] <http://www.lockss.org/about/principles/>

Author Biographies

Richard J. Solomon is a Visiting Scholar in the Moore School at the University of Pennsylvania, and chief scientist at Creative Technology, LLC, where he has been researching wave-based imaging sensors and human vision for US DoD and NASA for 15 years. Formerly, he was Associate Director of the Research Program on Communications Policy at the Massachusetts Institute of Technology, working on the MIT/Polaroid/Philips HDTV camera for NASA and DARPA, and on advanced telecommunications technologies.

Melitte Buchman is the Digital Content Manager at NYU’s Division of Libraries, responsible for digital imaging and preservation of video. Prior to NYU she worked at The New York Public Library both in their Digital Library Program, as Head of the Digital Imaging Unit, and in the Exhibition Department. She also serves on the board of Independent Media Art Preservation, Association of Moving Image Archivist’s Magnetic Tape Crisis Committee and the Center for Alternative Photography.

Clark E. Johnson has more than 60 years experience as a magnetics expert and physicist. At 3M he directed R&D on photoelectrically active materials for data recording and reading; optical analysis of retro-reflective media (e.g. “Scotchlite”); and advanced magnetic recording technologies. As an IEEE Fellow he has been an advisor to the US House of Representatives Science Committee, and as a consultant to US DoD on digital HDTV implementation. University of Minnesota, BS (Physics) and MSEE.

Eric Rosenthal, CEO/CTO of Creative Technology, LLC, is Adjunct Professor/Scientist in Residence at New York University’s Interactive Telecommunications Program, teaching Master’s classes in electronics and digital imaging. Over 40 years experience in electronics technologies including advanced, wave-based imaging for US DoD and NASA, low-cost spectrometric sensors, a novel 3D video system, and micro-miniature directional microphones. Formerly, VP Advanced Technology Research at Walt Disney Imagineering Research and Development, and General Manager Systems Engineering for the Disney/ABC network.

Dr. Jonathan M. Smith is the Olga and Alberico Pompa Professor of Engineering and Applied Science, and Computer and Information Science, at the University of Pennsylvania. Formerly a DARPA Program Manager (awarded OSD Medal for Exceptional Public Service), Bell Telephone Laboratories, and Bell Communications Research. An IEEE Fellow, and on the National Research Council Board on Army Science and Technology. Current research includes programmable network infrastructures, cognitive radios, and architectures for computer-augmented immune response.