ROTTEN PIXELS

ARTICLE

Jordan Craig

PHOTOGRAPHS

Collection of the CRP&A

RAILROAD HISTORIANS AND ARCHIVISTS are well aware that photographs fade and documents yellow with age, but likely few understand how digital files can decay over time as well. Even fewer know what to do about this threat lurking in their drives, discs, and other storage media. As the new digital projects coordinator for the CRP&A, I recently implemented systems to monitor our digitized photographs for this gradual corruption—what computer scientists call "bit rot"—and take restorative action when it occurs. Even if you are not a programmer, or perhaps recoil at the sight of the command line or mention of code, you can still use the principles behind these archival processes to protect digital assets of your own.

Tiny changes, massive problems

When viewing a digital photograph of a steam locomotive, you might see smoke and steel on your screen. But your computer sees something completely different: potentially millions of 1s and 0s. Every 1 and 0 is called a bit, which is the smallest, most basic building block of all digital information, including images. Despite their tiny size, changing just one of these bits from a 0 to a 1 or vice versa can have devastating consequences.

If you are unfamiliar with these 0s and 1s (known as binary code), it may seem logical for a single bit to correspond to a single pixel, but the reality can be much more complicated. This is especially true for a compressed image, like a JPG file, which is made smaller by removing redundant data. One bit change could make an image file completely unreadable, or even dramatically alter how the photograph appears. You can think of a bit flipping like a car derailing in a long freight train. That one car will probably drag the cars behind it off the tracks, creating catastrophic damage far beyond the initial point of failure. Similarly, just one flipped bit can set off a brutal chain of destruction in image files.

Archivist suggestion: *If you don't already, consider storing master copies of your photographs in uncompressed formats like TIF or camera RAW files. They are less susceptible to cascading corruption effects that can severely damage compressed images like JPG.*

Unless you intentionally flip bits around (as I did to create the examples on the facing page), these changes can result from technological problems or natural degradation. If we don't want our digital railroad photography turning into abstract art, we need to know why bit rot occurs, how to detect it (even within large image collections), and how to be prepared when it strikes.

Digital images are still physical

To understand why bit rot happens, we must remember that every digital file requires a tangible object to exist. This could be the hard disk inside a computer, a flash drive buried in a desk drawer, or a CD nestled in a dusty box somewhere. Even if you store your digital images "in the cloud," they are not invisible 0s and 1s ethereally swirling all around us. Rather, the data lives on machines owned by companies like Google or Dropbox. This matters because just like the film and paper we typically preserve in archives, electronic storage media can deteriorate and be destroyed.

Bit rot mostly occurs because the same environmental threats that damage analog photographs also attack the devices holding digital images. Physically, bits can take many forms, such as tiny magnetized areas on hard drives or microscopic pits on optical discs. Tangible 0s can gradually turn to 1s and 1s to 0s when altered by heat, humidity, debris, scratches, and other external forces—not to mention issues that uniquely affect electronics, such as magnetic interference or even cosmic radiation at high altitudes.

Archivist suggestion: *If possible, store digital media in a cool, dark place with stable humidity, but understand that proper storage only slows degradation. It will not stop it completely. Therefore, it is also a good practice to migrate digital files to new devices as older ones age.*

All of this might sound alarming. However, know that bit rot is an incredibly slow process. It usually takes several years to cause noticeable problems. This means that even if you are just learning about data integrity for the first time, it is probably not too late to take action and safeguard your photographs for the future. To do this, we first need to establish a baseline of health for collections and be able to recognize data decay when it happens.

Catching corruption

While you could periodically open every single image file you have and visually inspect it for traces of bit rot, that would be horribly tedious and inefficient.

Ron Hill's 1962 photograph of a Chicago North Shore & Milwaukee *Electroliner* train at the station in Milwaukee, Wisconsin, is the basis for these simulations of bit rot. The original JPG file contains 2,628,720 bits of information. Each of the sixteen corrupted versions of the file is the result of a *single* flipped bit. Hill-19-13-09


NORTH SHORE LINE

That is definitely not an option for us at the CRP&A, where we have hundreds of thousands of digitized preservation master images in our collections. So how can we detect this problem? Our most powerful tool is math: specifically, a hashing algorithm. And I promise that for our purposes, it is much less complicated than it sounds.

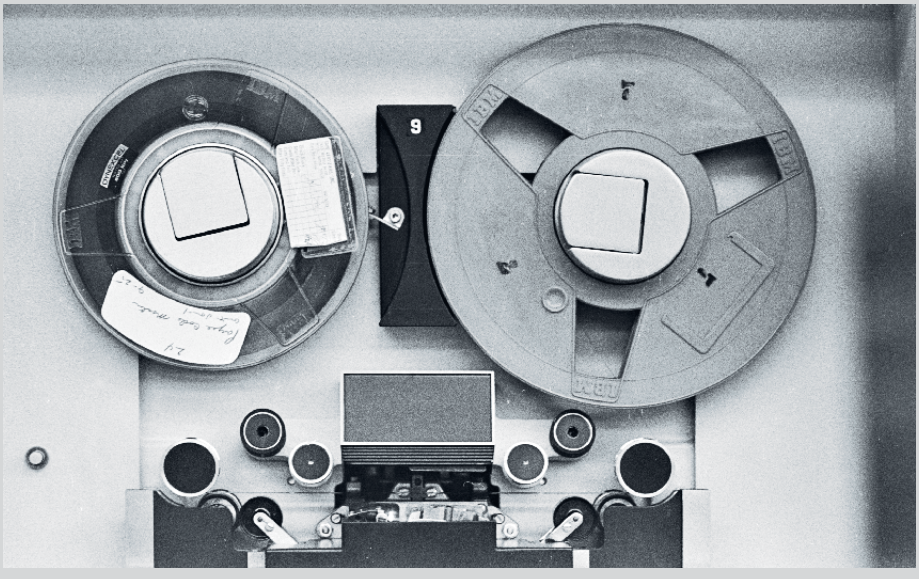
A hashing algorithm essentially takes all of the 1s and 0s that comprise an image file, runs them through some mathematical operations, and generates an alphanumeric string (i.e., a bunch of letters and numbers) to uniquely identify the image. This unique identifier is called a checksum. You can try this yourself by uploading an image to a tool like https://appdevtools.com/checksum-calculator, selecting an algorithm, and seeing the result.

The most crucial point to understand is that a hashing algorithm will always produce the same checksum when run on a healthy image file. When you run the same algorithm on an image months or years later, you should get identical results. If you do not, you know something has gone wrong. That something could be corrupted bits. Therefore, it is extremely important to create checksums for images while they are healthy, ideally soon after creation, and then periodically verify that those checksums haven't changed.

To generate and verify checksums for several files at once, I like to use the hashlib module for Python (a versatile programming language named after the British comedy troupe). If you are comfortable using command line interfaces like Windows PowerShell or the Mac Terminal, those have built-in hashing commands as well. However, if you prefer a no-code option with a graphical user interface, there are free utilities available. One example is QuickHash GUI (https://www.quickhash-gui.org/), which can process and compare entire folders of images with the click of a mouse button.

Archivist suggestion: *If you don't have existing checksums to verify your image files' integrity, but you have backup copies of your images, try creating checksums for both your orginal images and their backups, and then compare those. The corresponding originals and backups should have matching checksums. If they don't, at least one of the copies has likely been corrupted.*

Which algorithm you choose is completely up to you. Some, like MD5, are faster and produce shorter checksums, but are more likely to collide. This means that multiple files could produce the same checksum. Others, like SHA-256, are slower to run but generate longer, more secure checksums. For reference, we use SHA-512 at the CRP&A. This honestly might be more robust than most people need, but we are an archival organization. Regardless of which algorithm you select, just be consistent and take note of which algorithm you used. Also, make sure to store the file names and checksums in a nonproprietary format like CSV or TSV, so the integrity of your digital images can hopefully be verified in the years, decades, or even centuries to come.



The physical nature of digital information is especially apparent in early railroad computing, as seen in the Soo Line's IBM System/360 photographed by Wallace W. Abbey in 1969. Technicians stored data on objects that are typically associated with archives today, such as reels of magnetic tape (below) and paper-based punch cards (lower left of facing page). Abbey-07-059-27

Damage control

So what happens if checksums reveal the worst-case scenario? The bad news is, if you are unprepared, there's not much you can do about bit rot. The good news is that effective preparation is surprisingly straightforward. You may already be prepared without realizing it. This is because the easiest method of dealing with bit rot is simply having a healthy backup copy to restore from. Just overwrite the damaged image with its unchanged counterpart.

Archivist suggestion: *The National Digital Stewardship Alliance recommends keeping two to three distinct copies of archival data, preferably with at least one on a different media type and one in a separate geographic location.*

While manual restorations from backups work well for small personal collections, larger archives like ours

Left and lower left: Additional 1969 photographs of Soo Line's IBM System/360 by Wallace W. Abbey. Abbey-07-060-22 and Abbey-07-060-28

Below: Baltimore & Ohio's Honeywell Datamatic 1000, at Baltimore, Maryland, in 1959. The room-sized system used vacuum tubes, crystal diodes, and magnetic tapes. It could use a maximum of 100 tapes, providing a total storage capacity of 22.32 gigabytes. Fingernail-sized Micro SD cards can now hold more information, but the tapes are likely more stable. Photograph by David Mainey, Mainey-14-154-011

often require automated solutions. These include self-healing file systems that run checksum verification on files upon access and repair any damaged data using redundant storage. Self-healing can be compared to the process of solving a Sudoku puzzle. All of the bits are organized into sections following particular mathematical rules. Just like in Sudoku, the system can logically use those rules to determine when something doesn't add up, what the correct value should be, and fix it automatically.

At the CRP&A, we specifically use OpenZFS, a self-healing file system that handles this detection and correction process for our hundreds of thousands of photographs without requiring manual intervention for every single instance of corruption. Additionally, we maintain a separate record of SHA-512 checksums for our digitized images, providing an extra layer of verification beyond the file system's built-in protection. Better to be overly cautious and safe than sorry when it comes to our image collections.

Bit rot roulette

When I first started altering random 0s and 1s in a Ron Hill photograph to simulate bit rot for this article, some versions looked normal after several changes, while others became nearly unrecognizable after just one. I sent a few examples to Scott Lothes, our editor and executive director, and his response captures this phenomenon perfectly. After expressing some incredulity, Scott likened the unpredictability of bit rot to a game of roulette, where the stakes are image destruction. Do nothing, and you are placing a bet on that spinning wheel. We at the CRP&A are not gambling with railroad heritage. Are you? •

See more photographs from our collections in the gallery that begins on p. 62.

Opposite: Delaware & Hudson Railway workers prepare Baldwin "Shark" locomotive 1216 for new paint at the Colonie shops in Watervliet, New York, on September 10, 1974. Processing archivist Natalie Krecek has recently finished her work on the nearly 30,000 black-and-white negatives in the Shaughnessy Collection and is now starting on his color slides. Photograph by Jim Shaughnessy, Shaughnessy-N-DH-2112

Collection	Processing Status
Jim Shaughnessy	Negatives complete; slides in initial planning
John Gruber	Negatives complete; slides: fall 2025
Henry Posner III	Odyssey posting in progress
Steinheimer / Burman	In progress: 55% of slides complete
Karl Zimmermann	In progress: 25% of images onsite complete
Stan Kistler	Slides: 80% complete; negatives: fall 2025
Keith Bryant	Estimated start: late 2025

Railroad Heritage Visual Archive update

At our Monroe Street office in Madison, Natalie Krecek has finished processing Jim Shaughnessy's negatives. *Huzzah!* What a milestone! Processing as well as providing archival care for the Shaughnessy negatives, which needed individual resleeving for preservation reasons, has been a time-consuming endeavor. Over the five-year duration of this project, Natalie has single-handedly rehoused and described nearly 30,000 negatives, digitizing approximately 75% of the lot. She has completed a preliminary finding aid for the negatives, which you can find on our website, and she is now working on the much smaller slide series in the collection.

Reference and processing archivist Gil Taylor, who works at our South Park Street location, has a milestone of his own to celebrate. He's completed the selective digitization of Stan Kistler's slide series and is now recording and cleaning up the associated metadata. Gil estimates he's a little over halfway through this work. He plans to post selections from Kistler's color photography to Odyssey this fall.

Associate archivist Heather Sonntag has surpassed the half-way point with her processing work on the Richard Steinheimer and Shirley Burman Collection. She's currently working on a portion of the collection that she's nicknamed "the desert series," which depicts railroad stations and environments between Mojave, California, and Prescott, Arizona, including the Arizona Divide.

Beyond penning this article, Jordan has also made significant progress with processing the Karl Zimmermann Collection. Guided by Karl's own selections, Jordan has digitized 350 images shot been 1959 and 1970 that depict both domestic and international rail subjects. As this group has little written metadata, she has sent the digitized files to Karl for further description.

Finally, I've also carved out a little processing time for myself over the past month or so. I'm currently digitizing Richard Steinheimer's coverage of the Milwaukee Road's electric lines in the 1970s. Many of these images were included in Steinheimer's black-and-white masterpiece, *The Electric Way Across the Mountains*. The fun thing about this series is it includes several outtakes and alternative views that were not featured in the book. I hope admirers of *Electric Way* enjoy the additional images when we post them on Odyssey in the coming months!

—Adrienne Evans


1216