



EXPIRED

**DIGITAL INSURANCE
FOR INFORMATION AT RISK.**

**A STRATEGIC OVERVIEW
OF DIGITAL PRESERVATION.**

**H. Andrew Lawrence, Worldwide Product Marketing Manager
Document Imaging, Eastman Kodak Company**



HOW SOON WILL YOU GET YOUR DIGITAL WAKEUP CALL

The information contained by digital documents is at risk. The fact that a file, record, or image is backed up on tape or CD is no guarantee of permanence. The hardware and software required to read the digital code will almost certainly be obsolete within five years, while storage media are subject to abuse and aging. Kodak offers a digital preservation strategy that assures future access regardless of technological change, while avoiding migration.

...our digital documents are far more fragile than paper. In fact, the record of the entire present period of history is in jeopardy. The content and historical value of many government, organizational, legal, financial, and technical records, scientific databases, and personal documents may be irretrievably lost to future generations if we do not take steps to preserve them.

Jeff Rothenberg¹

IT'S THE YEAR 2025. DO YOU KNOW WHERE YOUR INFORMATION IS?

Imagine it's twenty-five years in the future. Your financial services firm is mired in a class-action suit that could bankrupt the company.

However, preliminary legal research points to the existence of evidence — an e-mail string created 25 years ago, complete with electronic signatures — that will clear the company of any negligence or wrong-doing.

As the legal team sorts through stacks of CDs in cartons brought up from the sub-basement of corporate headquarters, hope for a successful dismissal quickly turns to despair. Some CDs have delaminated. Testing reveals that the files on the intact CDs require application software that has been superseded by many generations. The media was not updated and migrated to a more current format. The evidence that is key to the company's defense cannot be accessed.

This scenario, admittedly an extreme example, reveals one of the major dilemmas for organizations in the digital age. Digital information is only as permanent as the hardware and software that give it intelligibility.

As technological evolution over even the past ten years has shown, there is little assurance that today's information will retain any more value than a 1989 VisiCalc file stored on an eight-inch floppy disk. Retaining digital information requires constant human attention.

File corruption

Format not

ALL? to read media.

DISAPPEARING ACTS.

The reports started appearing in the press in the late 1990's.

- Magnetic tape containing 1970's era satellite photo survey data of the Brazilian Amazon cannot be accessed to establish deforestation trends.
- Twenty percent of the data collected during the 1976 Viking Mars mission can no longer be read.
- In the State of Oregon, the primary database of people with disabilities vanished.
- Some POW and MIA records and casualty counts from the Vietnam War can no longer be read.
- At Pennsylvania State University, all but 14 of some 3,000 computer files containing student records and school history are no longer accessible.

Much more has probably been lost. The problem is that you don't know you have a gap in your digital records until you try — and fail — to retrieve something. By then it's too late.

[The cost of accessing and recovering digital information may far exceed the cost of dealing with the year 2000 computer problem.
National Science Foundation²

DIGITAL DOCUMENTS: A NEED FOR LONGEVITY.

Clearly, there are some classes of information we must keep for the very long term: patent records, FDA compliance filings, legal documents, records of business activities, and historical data. The growth of communications based on pictorial content (InfoImaging) driven by new channels, such as the Worldwide Web with electronic signatures, adds another layer of digital documents. Many of these merit preservation with additional complexity and expense.

All of this information is priceless when required:

- to research a customer database for a new product launch.
- to provide a fast, clear response to a subpoena.
- to settle a contractual dispute or claim by a policy holder.
- to meet the legal requirements for corporate record keeping.
- to find a suspect's earlier arrest photo among police records.

You may need to maintain accessibility for decades or a lifetime, through multiple changes in hardware, software, and media technology.

What are digital documents? Where do they come from?

Document images,
scanned from paper or faxed into your system.

Reports and statements,
generated by accounting, EDI, and e-commerce systems.

Word processing files, spreadsheet, and emails,
created in a collaborative or individual work environment.

InfoImaging content,
ranging from insurance claim photos and product Web pages to medical diagnostic images.

What could you lose?

A media problem or a hacker event could lead to the loss of money, your livelihood, and even your personal freedom.

Personal documents

- Family memories
- Proof of investment
- Retirement funds
- Insurance benefits

Business documents

- Corporate records
- Patent rights
- IRS penalties
- Research data
- Patient records
- Proof of transaction

ted.

recognized.

BY ALL MEANS, CRUISE THE INFORMATION SUPERHIGHWAY.

CAN A FACT BE “FOREVER” IN THE DIGITAL AGE?

Political, corporate, and personal history survives because we record facts about the present. But when the facts are committed to digital media, their lifetimes can be brief due to a variety of perils:

- mistakes in backups or storage purging routines.
- media damage due to material failure, aging, or a disaster.
- unavailability of functional hardware, software, or drives.

This last point is the most insidious, because obsolescence is the factor that is most beyond your control. Without the right technology, digital records, no matter how carefully managed, remain enigmatic strings of ones and zeros. If a single bit is out of place, the software’s ability to interpret the entire content and display or print it is lost. Compression routines and encryption only compound the problem. Without the guarantees of media permanence or backwards compatibility, “forever” can be as short as five years or even less.

Caught by a digital “gotcha.”

In the course of preparing this paper, the author found a Web abstract that referred to a study about digital loss, complete with a clickable link.

Surprise: “The requested URL could not be retrieved.” The report was no longer posted! Luckily, a colleague had downloaded and printed a copy.

Clearly, part of the problem is that there are few formal preservation policies or processes in place. And the drive to automate business in the name of customer service, cost, and profits is only intensifying. As more imaging, workflow, COLD, and ERP systems come on line, a growing volume of documents are being born digital at the desktop and on the Web, living digital, and, unfortunately, dying digital without ever making it to paper. Valuable business intelligence is, in effect, floating around unmanaged and unpreserved.

In one case involving FDA-mandated records of drug testing, blood-pressure numbers were randomly off by up to eight digits from those in original records following data transfer from UNIX platforms to Windows NT operating systems.
As reported in Business Week⁵

LEMMINGS MIGRATE, TOO.

A true preservation strategy must put planned-out business rules behind storage migration as opposed to simply moving information as capacities fill up. Even then, digital information is lost due to human error and the failure to make backups in the first place. This is compounded by media problems and technological obsolescence.

The practice of routinely copying, or “refreshing,” stored tapes to new tapes has been a familiar procedure in centralized, mainframe computing environments. It is labor intensive and expensive. But it works — as

long as the computer platform and software doesn’t change.

Migration strategies attempt to address the problem of technological changes. As drive and media platforms evolve, files are copied from one media to the next. If the application software changes without the guarantee of backwards compatibility, a conversion step must be undertaken to maintain access to the data.

Unfortunately, these procedures are error-prone. One analyst has compared them to the children’s game of “Telephone” in which a whispered phrase mutates as it passes around a circle. Just think of how a CEO’s off-the-cuff comment gets distorted as it passes from cubicle to cubicle.

With digital documents, the body text on a page may survive conversion from one word-processing program to another. However, formatting, headers, footers, and graphic objects are frequently scrambled or lost. Even something as simple as a font substitution can degrade the readability of the information.

Migration is also very expensive. In many instances it can cost up to 2.5 times the original cost of creating the information in the first place! It siphons away resources that could have been used for forward-looking, revenue-generating activities. When retention and back-file conversion compete with growth-oriented projects for budget, retention usually loses. For rarely retrieved documents, the benefits of online access just aren’t worth the monumental expense of digital migration.

for potholes.

...a time bomb whose full impact will register only in the future.

Donald Waters, Director,
Digital Library Federation⁴

WHAT'S NEEDED TO ACHIEVE DIGITAL PRESERVATION?

Kodak has studied the challenges associated with preserving digital documents as a natural extension of its 70-year history of offering document retention solutions. The company has identified the following issues as vital to a well-thought-out preservation strategy.

- Longevity – the ability to access and read digital documents in the future with virtually no degradation in information content, including formatting.
- Inter-operability – the ability of the digital preservation technology to work with existing and future IT systems to provide access to the preserved documents.
- Total Cost of Ownership – a lower total lifecycle cost viewed over the long term, including capital expenditures, storage media, operational expenses, maintenance contracts, and migration/conversion costs.
- Technology obsolescence protection – the ability to provide access to facts contained within digital documents through successive generations of new software and hardware.
- Backup and recovery support – the ability to support business functions, if necessary, using preserved documents.

Emerging digital preservation technology is offering a way to deliver permanence while addressing all of these issues. This is welcome news for document-intensive organizations with longer-term retention requirements such as insurance companies, financial institutions, health care centers, hospitals, government agencies, and public utilities.





IN THE LONG TERM, A SINGLE IMAGE IS WORTH A MEGABYTE OF DIGITAL.

DIGITAL-ANALOG-DIGITAL. KODAK'S TWO-WAY STREET TO DIGITAL PRESERVATION AND ACCESS.

The concept underlying Kodak's digital preservation system is diagrammed below.

Digital document files, including images, are assembled from the host system and passed to a digital-to-analog processing step. They are then output to a permanent storage media in analog format (images), with indexing. At any point in the future, these images may be digitized and delivered back to the host system.

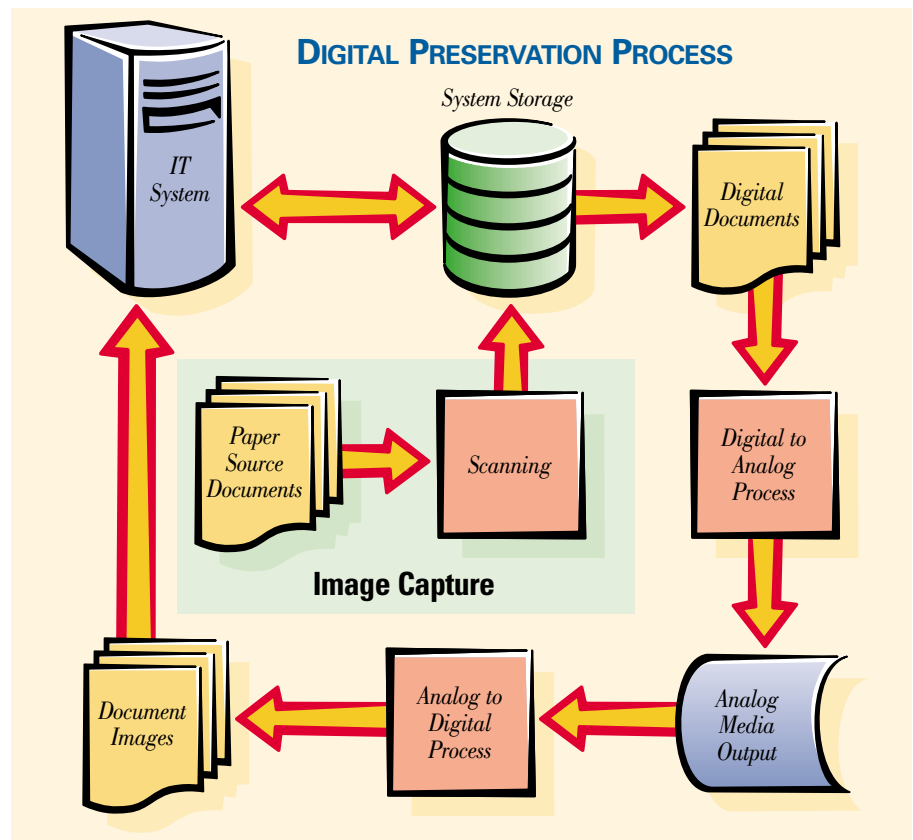
As indicated in the flowchart, image capture from paper is easily integrated into this process. This supports operations that include imaging for document management or automated workflow. In addition, it provides a digital "on-ramp" for

compact, organized retention, storage, and retrieval of paper documents.

With the proper technology, digital preservation can take place as a background process, requiring little, if any, attention. Digital files can be allowed to expire naturally and then purged from the system. Organizations may choose to outsource the actual output, physical storage, and retrieval to a third-party specialist, such as an imaging service bureau, an Application Services Provider, or an online service.

Individuals and businesses that create and drive our economy, our government, and our educational and research institutions face a similar problem of Titanic proportions.

The Titanic 2020 Problem
Censa³



1000 1010

ENABLING DIGITAL PRESERVATION WITH INTEGRATED IMAGING FROM KODAK.

All of the elements are in place in the form of commercially available Integrated Imaging products that bridge the gap between digital and analog. Many Kodak customers have already implemented some form of digital preservation. These solutions include:

- Software for queuing, managing, and converting digital documents downloaded from the host system (electronic systems).
- Output hardware that uses exclusive LED array imaging technology to create very reduced analog images at high speed.
- Archival Storage Media that is ANSI/ISO-certified for a 500-year life expectancy when properly processed and stored.
- Analog media scanning hardware that retrieves, digitizes, and returns digital documents to host systems automatically.
- Flexible document scanners for capturing images from paper source documents.
- Necessary service and support provided by Kodak, its reseller channel, and authorized Document Data Conversion Centers.

Kodak is involved in ongoing research and development to broaden the applications for its current and future Integrated Imaging technologies. Together with strategic partners and customers, the company is working to assure we will be able to preserve those digital pieces of the present we want to keep.

The historical significance of many of our digital documents — which we may not consider important enough to warrant saving — may become apparent only long after they have become unreadable.

Jeff Rothenberg⁶

THE TIME IS NOW FOR DIGITAL INSURANCE.

Clearly there's a need for simple, affordable strategies that assure the long-term survivability of information. The risks of loss are immense. Hidden in the flood of information could be statistics that identify cancer-causing environmental agents, priceless firsthand accounts of history as it happened, and the owner data for the next product safety recall. Integrated Imaging technology from Kodak offers a starting point and a springboard into the future.

Got analog? It's all around us.

Examples of analog storage:

- The Dead Sea Scrolls
- The Rosetta Stone
- Shakespeare's Folios
- The Declaration of Independence
- Daguerreotypes of Lincoln
- Microform fiche and film
- Any paper book
- A stock quote on your WAP cell phone
- The e-mail you printed out this morning and put in a file folder.

Digital preservation implementations.

- Vital Records and Archives
County Governments
- Lotus Notes and MS Office Attachments
Singapore Housing Board
- Federal Records
Malaysia Registry of Company
- Census Forms
United Kingdom Office of National Statistics

More detailed product information and customer profiles are available at www.kodak.com/go/Integratedimaging

SOURCES

1. Rothenberg, Jeff. "Ensuring the Longevity of Digital Information," published by RAND, February 22, 1999
2. Cited in "The day that the data die," *Rochester Democrat & Chronicle*, page 3E, October 31, 1999
3. "Titanic 2020," issued by the Collaborative Electronic Notebook Systems Association (CENSA), <http://www.censa.org/Titanic2020.html>, January 2000
4. Cited in "Whoops, there goes another CD-ROM," *U.S. News & World Report*, February 16, 1998
5. "From Digits to Dust," *Business Week*, April 20, 1998
6. Rothenberg, Jeff, op. cit.

