Project 34 - Analog Preservation of Paper and E-Documents

Robert Breslawski
Eastman Kodak Company
Rochester, New York, USA

Objectives:

Share KODAK's vision for Reference Archiving
Current commercial techniques
Quality Considerations
Related Software Techniques

Agenda:

Modern document management dynamics.
Digital to Film"
Current Technologies
Quality Control. A System Approach

Today's document management dynamics include rapid access to all types of information including computer data, paper documents, web pages and more. The rapid transition to the eDocument has enabled easy access, transmission and duplication. The acceptance of digital documents and contracts has become widespread and official sanctioned with the signing of digital signature legislation by President Clinton. What it has not provided is a cost effective preservation of the human experience related to the digital information accessed. The challenges of eDocuments and eRecords are the issues of authenticity, integrity, trustworthiness and sustainability. The latter becomes a serious issue when timeframes longer than five year are considered since migration and technology obsolescence become major factors.

Kodak's vision, the "Reference Archive", utilizing tried and true film imaging based technology along with the digital partnership can provide the answer to those issues. Capturing digital data and scanned in documents within moments of creation (finalization) in analog form on LE 500 year media fits the bill for authenticity, trustworthiness, integrity and sustainability. "Reference Archive" is a concept, not a product. The products or services that provide a "Reference Archive" come from many sources. The concept of preserving digital information on film seems to be an acceptable premise, where it breaks down frequently is in the implementation. This paper describes the currently available equipment, software and services that can provide users with a "Reference Archive" and also giving some guidance on how to handle grayscale images and related quality control issues.

So, once you have committed to utilizing a "Reference Archive (RA), how does one go about setting one up and using it. The answer is always of course what do you want to accomplish with it? Do you want to archive everything or a subset of documents? How do you want to organize and retrieve the documents? It's a little like taxes, up front organization and planning goes a long way.

The following should be considered in pre planning:

What types of "documents" will I be writing out to film?
- Scanned in images only.
- Desktop, web pages and other digitally created documents only.
- Some combination of the above.

How do I want to organize them on film?
- Format of the images
- Image Quality considerations – text size and type, not document size
- Image Orientation – cine or comic, or both
- Indexing schema – single level, multi-level

How will I be collecting the images?
- Scanner – resolution, thresholding, contrast, deskew, "Perfect Page" considerations.
- Digital conversion from desktop applications such as Microsoft Office or PDF.

How will I combine the appropriate images?
- Appropriate software to be chosen that works with my document management application

How will I record the images on film?
- Purchase my own equipment
- Outsource the filming operation to a service bureau.
  - If so, how do I get the documents (files) to the service bureau?
- Determine how to process the film, in-house, or service bureau.

What Life Expectancy (LE) do I need for my records?
- Silver gelatin is classified as LE 500 or

Where do I store the "Reference Archive" tapes once written?
- Considerations for self storage would be how to maintain environmental conditions as described in appropriate standards such as ISO 18911.

How will I utilize my reference archive?
- Considerations include physical handling of the media,

Scanning and transmission of the images to the user.
This can be done internally, or with an external service.

The following is a list of commercially available capture opportunities:

Kodak i9600 Series Writers
LED Technology Bitonal TIFF images to 16mm Microfilm
Conversion Rate of up to 400 Images/per minute
Industry standard formats.
Services provided by certified Document Conversion Centers
Writes at about 8000 dpi resolution.

Image Graphics Inc. (IGI)
eBeam technology
Multiple media formats
Grayscale or bitonal
Imaging Services

Microbox AG (Germany)
Laser Technology Bitonal to 35 mm
Specialize in Engineering Drawings
Table top size writer 2 frames per minute
LE 500 media

Quintek
Bitonal Laser Technology to 35 mm dry technology (LE 100)

Tameran
Archive Services – 35 mm film recording
Aperture Card plotters and supplies

Global Information Distributors AG (Germany)
All points addressable 105 mm fiche production systems

ANACOMP Inc.
Print and Imaging Services – 105 and 16 mm dry and silver-gelatin based films.

Image Recording

When choosing the method to write images, consideration should be given to what types of images and what the capabilities of the writer are. In some cases, such as with the IGI full 8 bit grayscale imaging is possible (but perhaps more expensive) than when written with the KODAK i9620 series bitonal writers. Don't forget, just because and image is "binary" or bitonal, it does not mean that the image cannot be half-toned and produced on film in such a way as to look like grayscale.

While there are multiple sources of writing images to film there are also considerations of what types of documents are to be written. If only scanned in documents in bitonal format are to be written, then scanner capture software such as KODAK Capture Software and Kodak Electronic Microimager Software may suffice. Others such as Kofax Ascent Capture can also facilitate writing images from both document scanners and digitally borne documents to film along with output of document indexes for later search and retrieval. The list of software providers that do this type of service include but may not be limited to:

- Kofax - Ascent Scribe
- Captiva/ActionPoint - InputAccel
- BIS - Masterscan 2000
- AMCAD - Land Information Software
- Results Engineering - OnBase Systems
- eiStream – KoVIS
- Filenet - Panagon Pro Svcs
- Image Solutions – PDFArchive
- Digitech - Paperflow
- Software Finesse – Flexsuite
- Hyland Software

Also, one has to consider that these image recording devices must rasterize the image prior to writing to film. For instance the Kodak i9600 Series writers will handle TIFF Group IV and or uncompressed bitonal tiff. Conversion utilities like Infotek Reformat are available to do such tasks as compression, rotation and annotation prior to the images being sent to a writer. It will also convert single to multi-page or multi-page to single page format depending on how the user wants the information to be stored on film. Software from ImageMaker.com can utilize a print driver to create writer compatible image files very effectively. These utilities may be provided as either full working copies or demo copies by the writer manufacturers.

Image Retrieval

Likewise, when retrieving documents from film and converting back to digital images, consideration must be given to what the output file format or rendered image is to look like. Analog to digital converters are commercially available from at least KODAK, Sunrise Imaging, Meckel Scanners, Wickes and Wilson Limited and many suppliers to who sell these and possibly other products as well. Time and space do not permit going into detail here but one needs to consider the needs of the user are. For instance, do you have legacy microfilms without image marks (indexed for retrieval), films with differing reduction ratios and orientations? The myriad of scanners and options for each requires study. For item retrieval and low to mid volume production retrievals, the Kodak i7300 scanner offering low and high resolution, up to 25 images per minute, grayscale and bitonal modes offers great flexibility for 16 mm applications. KODAK also supports a range of combination scanners called KODAK DSV 2400 and 3000 units that provide 16, 35 and 105 mm capability depending on options chosen. NextScan Eclipse scanners featuring high speed (up to 300 pages per minute) scan both 16 and 35 mm films. Higher volume scanners may be pricier but for large volume conversions may be a good choice. Be aware of the flexibility of interchanging lenses, film format platforms and

the like when choosing between scanner systems. The KODAK i7300 for instance features all choices in software from the PC workstation. Again, it is a similar situation with the writer, there are multiple providers of software that will allow reintroduction of film images to the digital world. A list of potential software providers who provide interface software complementing films scanners is as follows:

- eiStream - KoVIS
- K2 Microsystems - DistriNet
- ISSI - LandView
- PegaSystems – Reelay Trace
- MSI – DIDS for i7300
- Sterling Commerce - Vector Microimage
- VersaIMAGE- VersaIMAGE
- Hyland Software

Please note, this is not all there are, and these are certified to work with the KODAK i7300 Microimage Scanner.

Systems Quality Control

Lastly but not least, please consider a full systems approach to Quality Control. If one analyzes the components of a valuable RA, one will find high quality images being returned from high quality film. High quality images can lead to very high OCR efficiencies, thus allowing documents stored in analog format to be returned to being processable documents.

My recommendation would be to start quality control with the original document characteristics itself before deciding what type of film recording to use, meaning what are the document sizes and even more importantly how large is the print. For instance, it makes no difference if the document size is 8.5" x 11" or an E size engineering drawing if the text size is 4 point font, then a lower reduction ratio (scaling in the case of digital) is required. One should also determine what scanning or rasterizing resolution to use by considering font size and the number of available pixels that will be used in the writing device. For instance, if you have 4 to 6 point font, many elements of the characters will be single pixel wide. When this is recorded to film, it maybe necessary to thicken up the fonts so that loss does not occur during writing.

The following list is not all inclusive but at least a good start for consideration of requirements:

Assess document input quality

- Paper weight
- Font size
- Document contrast
- Determine scanning resolution based on recommendations in ANSI/AIIM TR 26 1993, Resolution As It Relates to Photographic & Electronic Imaging
- Consider what the writer specifications. It is typically optimal to scale down, not up so the document scanning resolution should adjusted to match the output of the writer if possible.
- As a practical example, the Kodak i9620 writer can put 3888 pixels across 16 mm film, If you scan an 11" document at 300 dpi, then you will have 3300 pixels and when written unscaled this image will fit easily on the film and will have the best image quality. Write pixel for pixel whenever possible.

Quality control the writing device

- Unlike traditional optical imaging systems where quality of the filming unit cannot easily be separated from the input quality of the test targets, with digital writing, the writer/film combination can be independently assessed using either a digital (software) target provided by the manufacturer of the writer, or and industry standard target such as that described in ISO 14648 -Micrographics - Quality control of COM recorders that generate images using a single internal display system - Part 2: Method of use
- KODAK for example, supplies a target specific to its i9600 writers which evaluates all the pixels located on the LED bar. Similar digital targets are available from other suppliers as well. The key is to write these targets without scaling so that the full resolution of the writer can be assessed.
- Background density measurements should be coordinated with image quality characteristics. Depending on the type of film and recorder used, some systems will produce negative images with a background of around 1.0, whereas some recorders, typically those used in traditional computer output microfilm will be around 2.0. There are also positive appearing systems used which are handled differently, though writing a dark patch at what has been determined to be optimum character exposure as a daily control tool is a good starting point.

Quality control for the combined scanner/writer/film system tends to be a bit more complex.

- A good opportunity is to utilize a "practical quality" target such as that provided by AIIM International call /AIIM Scanner Test Chart #2. This target features various text fonts and sizes that can be compared to your own text and used to verify that what is on the film is readable. These targets can then of course be scanned back from the film during retrieval to verify that the entire system is indeed reproducing what is desired.
- These targets can be utilized in document scanning and film writing for daily quality control of production. More information can be found on these targets and other issues related to scanning in ANSI/AIIM MS 44 - Recommended Practice for Quality Control of Image Scanners.

Quality control for film processing

- Daily process control procedures are an expectation of any quality production operation using film.

- Consult ANSI/AIIM MS 23 Practice for Operational Procedures/Inspection & Quality Control of First-Generation Silver-Gelatin Microfilm of Documents or KODAK Pamphlet D 17 Control Procedures in Microfilm Processing.

Quality Control – evaluation after imaging

- Every user and laboratory needs to determine what level and type of inspection is required after the imaging production is done.
- Evaluation of resolution targets, quality targets and background density on each roll of processed film is a good starting point.
- Guidance is given again in ANSI/AIIM MS 23 Practice for Operational Procedures/Inspection & Quality Control of First-Generation Silver-Gelatin Microfilm of Documents and also ANSI/AIIM TR 34 - Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) & Micrographics Systems.

Quality Control – Retrieval

- The user needs to be aware that digital aliasing can become an issue when trying to read standardized resolution targets when converting from paper to digital, back to analog (film) and from analog back to digital (scanned) images.
- Ultimately, it is the final output quality that matters.
- For full systems control, a reference document should be utilized. This can be either a paper or even digital document, as long as you are comparing correctly. For instance, compare a paper original to a printed copy, or compare a digital image/document to the digital rendition on the same viewer.

General Image Quality Issues

Within the Quality Control portion above we discussed how to start and maintain a certain level of quality. What if you are not happy with the input and output of the system? For instance, the images may look acceptable on a display screen with a resolution of 72 dpi, but when written to film and rendered back, they may have lost something. Or, what if your documents have low contrast or pictorial subject matter? You can consider separating the text parts of compound documents from the pictorial part but you may not have to. We have done extensive work looking at pre-sharpening prior to writing out such that images and text are enhanced to improve image quality. Half-tone images (like newspapers for example) can be utilized on bitonal writers and achieve significantly smaller file size as compared to full grayscale images.

Weblink References:

http://www.kodak.com/global/en/business/index.jhtml?pq-path=2/782
http://www.tameran.com/
http://www.nextscan.com/
http://www.igraph.com/
http://www.crowleymicrographics.com/
http://www.sunriseimaging.com/
http://www.nextscan.com/
http://www.aiim.org/ (for ANSI and AIIM standards issues)

Biography

Robert Breslawski, Technical Associate, Eastman Kodak Company Research Laboratories

Since joining the Eastman Kodak Company in 1973, Robert Breslawski has acquired extensive experience in the development and worldwide promotion of enabling technologies for digital image reduction to and from hardcopy via Reference Archive systems.

Recently his work was key to the development and commercialization of the Kodak ImageLink i9620 series writers and the formulation of Kodak Reference Archive Media solutions.

Robert Breslawski, as a Chair of several committees, is a leader in standards development activities with the Association for Information and Image Management (AIIM) and the International Organization for Standardization (ISO) Technical Committee 171 Document Imaging Applications. (585) 477-6771; Email: Robert.Breslawski@kodak.com