

Kirtas Technologies

Automated Book Digitization
and Image Processing



Moving Knowledge from Books to Bytes.

Michael Maxwell

Bound Book Digitization System and Processing

Knowledge Trapped

Until recently, books and manuscripts have been the primary media used by researchers, scientists, historians, philosophers, and authors to record their works for reference or reading by others. The book format has been used for more than a thousand years. The number of book volumes in existence today is enormous and the amount of knowledge they contain is vast and invaluable to all of mankind. Yet, access to most of this information is highly constrained.



The existence of many of these books is often unknown by those that could benefit from them. Historians could expand their research, scientists could advance their work to the benefit of millions. There would be a better understanding of cultures that would advance all societies. The knowledge economy would accelerate if only more people knew of the book's existence, its general content, and had easy access to it.

Easy and quick access to the physical books is a challenge even if its existence is known. It is often impossible to access them at the physical level. Some books cannot be viewed because of the possibility of damage from use. Locked away and rarely viewed books are located around the world in over 200,000 public and private libraries, museums, archives and research facilities. It is an issue and a problem that was foreseen centuries ago....

"...let us save what remains: not by vaults and locks which fence them from the public eye, but by such a multiplication of copies, as shall place them beyond the reach of accident."

*Thomas Jefferson
February 18, 1791*

Knowledge At Risk

The books on the shelves in the great stone buildings are at risk for loss for many reasons. Some of these books are being lost through natural aging due to the acid in the paper. Some are lost because of public use, abuse and theft, and many are lost because of natural disasters like hurricanes, earthquakes, or tsunamis. The political unrest around the world is another condition that has put these valuable books at risk from terrorist acts. How unfortunate to lose these books, often one-of-a-kind, and the knowledge they contain.

While the physical loss cannot always be prevented, protecting and preserving the knowledge in the books can be achieved as part of the book digitization project.



Bound Book Digitization System and Processing

Trends

There are several trends that have inter-relationships and can positively effect the initiatives for book digitization. All of them are advancing independently yet their synergy offers the potential to release and make available all of the world's knowledge!

Globalization - The country's and culture's of the world are becoming less and less isolated. Changes in manufacturing and services delivery are dramatically different then just a decade ago. The economy is becoming more interdependent among countries. Virtual services are available around the world 24 hours a day. No longer is any country totally self-sufficient.



Knowledge - The Knowledge Economy is growing rapidly and those that can access that knowledge will be more successful in their research because they can more easily build on prior works. Others will be more competitive because they know more or can access more knowledge. The knowledge work-force is growing every year and represents a sizable part of the work force.



Digitization - With all the growth in the digital world, more the 85% of the total information is still stored in paper form. Now we can easily and inexpensively digitize that information to make it more readily available for others to access. Digitization also serves as a means to preserve the content of the books. Digitization technology continues to increase in value.



Internet - By sharing the information, the gap between developed and under-developed nations can be narrowed. Cultures will have a better understanding of one another. This technology has been rapidly adopted and allows us to go almost anywhere in the world in seconds. This conduit has expanded from its early government and research use to commercial and personal use that few could imagine 10 years ago. Investments continue to enhance and upgrade it and Internet 2 will be upon us shortly.



Book digitization needs to be done to achieve two important goals.

- release the knowledge from the shelves to help mankind
- preserve the content of the information for future access.

The digitization has to be done rapidly because more people are growing up believing that if something isn't on-line, it doesn't exist. If we don't digitize the rare books soon, we may never have the opportunity if they are lost to decay or disaster.

Objective of Digital Libraries

Having digital libraries that contain catalogued and indexed books and documents makes the library's content accessible to anyone with internet access. Electronically searching through catalogues to find sources and references of the information desired is efficient and easily linked with other library's catalogues. Making the world's knowledge globally available will add to the educational, economical and scientific advancements to better all people.

Bound Book Digitization System and Processing

Book scanning offers a key part of the solution for releasing the knowledge in the books to the people of the world. Using proven technologies available today, books can be digitized and kept in formats that are easily accessed and read on PCs with internet connectivity. The word content of many of the books can be processed through Optical Character Recognition software (OCR) to create computer based text files that can be loaded into databases. Now the content of the books can be found through word or phrase searching engines provided on the internet or other private programs. This is in addition to the traditional catalogue information that is typically associated with a book.

The image files can be easily produced and kept in multiple file formats to meet different needs. For example, the files can be stored as TIFF with no-loss compression applied to assure all the detailed information is retained; or they can be converted to a JPEG 2000, or PDF or compressed with other high compression algorithms to allow the files to move over networks; or optimized for internet viewing. Imagine for a moment how wonderful it would be for you to do your research in Paris and easily locate source information in Seattle or Cairo, to view the full text and all the graphics associated as if the original book was right in front of you!

Preservation of Knowledge

Physical preservation is one activity to keep knowledge from being lost. Preservation allows the retention of the original material, which will assure future access. The efforts to preserve the physical book are extensive and costly. By preserving the physical book, the content is also preserved. When books are at the stage of needing to be preserved, they are removed from circulation, which makes the content inaccessible. After the preservation is completed, many books have less accessibility in an attempt to prolong their existence because they remain fragile and easily damaged. So while the physical book is preserved, the content is lost or less accessible.

Book scanning facilitates the preservation of the knowledge content within the book (but not the preservation of the physical book). Regardless of whether the book is returned to circulation or not, the content remains accessible. The book is on the server instead of the shelf. Having multiple copies of the file on backup media will increase the likelihood of the file surviving disasters.

Preservation of the digital file is not easy or inexpensive. Technology obsolescence will occur and the digital files must be moved forward to the new platforms and technologies of the future. Selecting a mainstream format is important because those formats will be sustained longer; and investments to migrate those formats forward will be larger and achieved earlier due to market demand. The "LOCKSS"¹ principle is simple, and effective, and works easily with electronic files.

¹ Lots Of Copies Keeps Stuff Safe

Bound Book Digitization System and Processing

Quality and Standards

As with the use of any information tool, there must be standards established to provide for quality viewing just as there is for interconnectivity. If the quality is poor and viewers cannot obtain the information they desire, then the effort and investment to digitize was wasted. If the images are of poor quality, have obstructions, lack text clarity, or are missing pages, the users will quickly abandon the digital source. An internet visit to some sources of digitally scanned books will reveal the quality level varies widely.

Over time, the standards of quality will evolve and improve from what is the average on the internet today. As the quality standards increase, those that are marginal will be abandoned or will have to be rescanned to remain “competitive” with other sources. Poor quality images will reflect negatively on the owner.

A primary tenet of book scanning is to scan the book once and have a level of quality that will satisfy the viewers. Images must be of very high quality not only for the black and white text, but also for the color graphics content. Some goals that the project should define and achieve are:

- final image set is very high quality that would be equivalent to a file used for print-on-demand operations by a publisher to produce a book to sell to a customer, and derivatives for easy web viewing are a secondary consideration
- flexible enough to handle a wide range of book quality where the pages may be of different weights, color backgrounds, torn, lose or unintentionally covered up, or pages may be stuck together
- the electronic book file would be formatted to industry standards with technical and structural metadata provided to assure long term preservation of the file
- the scanning and processing will include the metadata for indexing the book in a database repository, aligned with the standards in existence e.g. Dublin Core +2

The file formats chosen should be industry standards and not proprietary. The readers for industry standard formats are readily available and easily installed on users computers. Proprietary formats may prevent users from viewing the images. If access becomes difficult, the files are essentially “lost”.

Additional considerations of the system should include:

- OCR processing leading to full text indexing for ease of content searching.
- Optimized file size to reduce the storage and network burden.

The primary applications are:

- Special collections
- Rare books and first editions
- Brittle page or acidified paper book collections
- Original manuscripts
- Musical compositions
- Land ownership
- Genealogically related records
- Scientific research documents

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Kirtas APT BookScan Design Criteria

With these known needs, a product idea was born in Xerox at their Palo Alto Research Center in the 1990's. People with a vision realized the future need to convert books to digital formats at a quality level and productivity rate that exceeded anything in the market at that time. The project team established some key tenets to guide them in their engineering and design efforts.

The engineering design objectives were:

- Book Safety While Scanning
 - Create a non-stressful position for the binding while scanning occurs
 - Create minimal stress on lifting and turning pages
 - Create as flat a page as possible for imaging
- Superior Image Quality
 - Capture pages in full color for accurate representation
 - Capture pages with the best Modulation Transfer Function available
 - Capture book with full integrity and no page loss
 - Process images to high quality industry standard formats
- Maximize Capture Speed
 - Minimize operator intervention
 - Use high speed capture technology
 - Use robotics for controllable, consistent performance

In late 2003 the first book-safe, high image quality, automated book scanner was shipped by Kirtas. Since then, nearly 50 units have been installed in locations around the world in libraries and service support organizations, as well as government and corporate departments to scan books in a safe, productive, non-destructive manner.

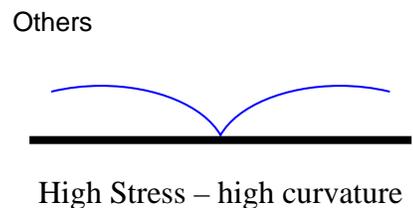
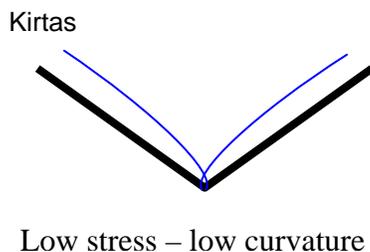
Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Non-stressful Binding Position

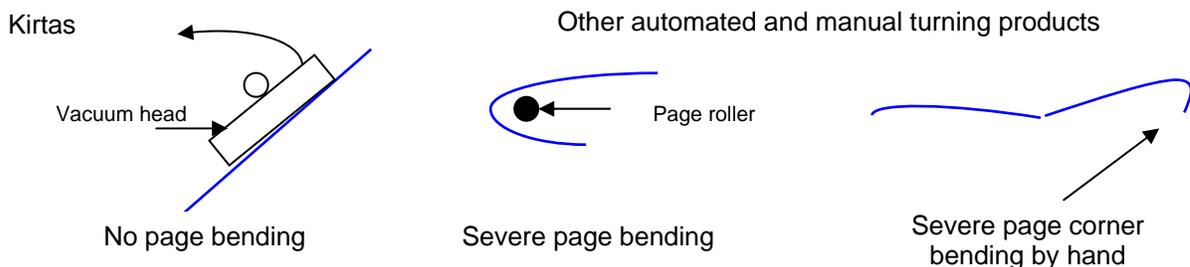
The engineering design of the APT BookScanner allows for non-destructive automated book scanning and minimizes the stress applied to the binding and the page turning. Kirtas's two patented technologies are the Smart Cradle™ and the Sure Turn Page Turner™.

The Smart Cradle holds the book open at a non-stressful 110-degree angle. As the pages are turned, the Smart Cradle angulations adjust to keep the book's spine in a non-stressful position while maintaining the proper photographic position. Having the book lay flat at a 180 degree angle will cause significant stress on all bindings whether new or old. The positioning also increases the curvature effect of the pages along the spine.



Minimal Page Turning Stress

The SureTurn Page Turner™ uses patented vacuum technology and directed airflow to gently separate and select the next page for turning by the robotic arm. The page separation is reliable because a light flow of air across the edge of the pages causes a lifting and separation of pages. The vacuum technology of the robotic head and the dynamics of airflow over the surfaces of the pages augment separation. The page is turned gently on the binding edge. Other systems roll the page in the tight circumference causing extreme bending of the page that will damage or destroy brittle and or older pages.



Robotic Reliability and Page Turning

The robotic arm technology is under system control and will repeat its function in the same way and at the same rate every time. The vacuum head lifts a large area of the page for turning in a gentle manner, not just the corner of a page. The arm turns the page consistently with the same force every time. Manual page turning for long periods will result in the operator varying greatly the force used and on what part of the page. This will eventually result in operator errors, including page tearing or other damage.

The robotic arm will be gentler on the book pages than the human hand for turning pages, and the arm will perform at a rate significantly faster than a human can achieve and sustain.

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Flat Page Positioning

In order to have the best image quality and avoid the curvature issues at the center spine and the edges, several systems flatten the pages in an attempt to get a flat plane for the scanner. Many use a platen of some shape to flatten the pages. The downside is the platen places excessive stress on the binding and pages as it attempts to flatten the book. In the end, some curvature is still present. Some systems use a “tiered copy-bed” to overcome the curvature. While it helps to get a flatter plane, some curvature remains and there is excessive stress put on the binding across the tiers.

This engineering approach is trying to take one scan or two pages to be as fast as possible. The result is a continuous compromise on image quality and potential damage to the binding and pages.

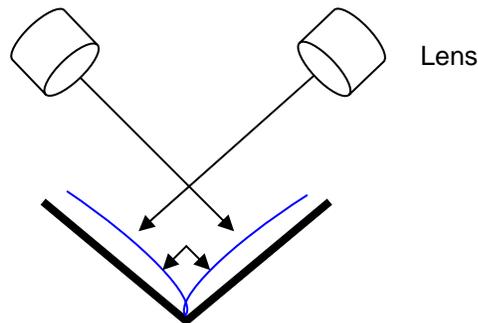
Others' Design



High stress on pages and binding – Curvature Still Present

The flatter the page is the better for imaging. With the Kirtas design, each page is intentionally scanned separately. Just by having the book in the cradle, most of the curvature problem is removed. The two clamps, positioned at the edge of the page near the binding, automatically and gently aid in page flattening after each page is turned. These clamps apply only a very gentle pressure to the page without creating stress on the binding. The clamps are easily positioned for varying book sizes and only touch the edge of the page with a downward contact. The depth of focus of the camera lens overcomes the slight remaining curvature (usually less than 0.5 inch).

Kirtas' Design



Low stress on page and binding – Curvature minimized

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Manual Only Scanning

Manually turning pages for bound books is the most common approach and might be a good economical choice for projects with a few books. However, for projects with medium to large quantities of books to convert, the manual process will be slow and have a higher incident of error. Operator fatigue will eventually lead to errors and page damage. A person turning pages will:

- use an inconsistent amount of force to turn a page resulting in tearing a page
- have variations in speed for turning each page affecting productivity
- have a hand or fingers in the capture image area blocking content
- have fatigue or loss of concentration resulting in multiple-page turns, double page exposures, speed loss

Beyond these, the time needed to complete a large project with manual page turning could take decades. The time to complete could take longer than the life of the scanner! Trying to scale up with multiple manual scanners and operators introduces additional points of variation in the cameras and the operators that will increase errors, and the Quality Control labor efforts.

Fully Automated Solutions

A few fully automated solutions are available that claim to preclude the need for operator attendance. "Put the book in, close the lid and come back when completed." While this may work for a small percentage of the book population, anyone that is knowledgeable about books knows that they can vary substantially. For example:

- bindings are stiff when new or unused, looser when highly used
- some pages may have corner folds
- pages will have different curl characteristics
- paper weight can vary widely
- pages can be stuck together
- loose items in the book that will block the capture of the text or picture behind it

All these variations will create different page turning requirements. Is it reasonable to think that one setting, unattended will accurately scan a book from start to end without any operator attendance or intervention and overcome these? With a closed system, the scanning operation could have errors that go undetected requiring full rescanning of the book or worse, cause damage to a book.

Manual Intervention

Because of these variables, it is beneficial to have a system that will allow the operator to have easy access for intervention to provide assistance to overcome these variables when necessary. The result will be a higher productivity rate with a high quality achievement in the capture scanning process.

There are three scenarios possible

- the operator can effect the assistance without stopping the robotic operation
- the operator can quickly and easily stop and restart the robotic function
- the operator can manually turn pages for short runs as needed

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

For example:

- Stiff bindings: some books will not open easily and page curl can be excessive resulting in poor image quality. the operator can help flatten pages
- Pages with tears: the operator can quickly pause the robotic arm and manually turn the pages so no more damage occurs, and then restart the arm.
- Loose pages: occasionally there are loose pages and the operator can stop the robotic arm and manually turn the pages; then restart once past the loose pages.
- Loose items: the operator can pause the operation, remove the note or debris and quickly restart the operation.

With the Kirtas design, the system will pace the operator. This is significantly different then the manual scanners. With the Kirtas design, the operator can intervene and assist the scanner without slowing down productivity – no other system offers this. With the Kirtas design, one can change between manual and automated scanning with just a touch of a button – no other system offers this. The result is the ability to achieve very high productivity across a wide range of books with different characteristics, yet maintaining the best image quality possible.

The speed of the robotic arm is suitable for a wide range of book variations, yet it is controllable to meet the characteristics of the unusual book.

Automated High Production Scanning

In order to achieve the highest productive throughput, the system must be indifferent to common variables of scanning. Typically, other systems will slow down significantly as these variables occur. For example, the table below illustrates a comparison between the Kirtas scanner with another scanning system. The number in the column represents the real percentage of throughput achieved compared to the manufacturer's quoted rate for best case scenario. Most manufacturers state the scan rate at 200 dpi, and letter/A4 size documents. Multiply that rate by the number the column to learn what the real scanning productivity number is. Please note that Best Case rate is not the same for each scanner. If Best Rate for Kirtas is 800 pages per hour and the Other scanner is 1000pph, then for 300 dpi scanning the Kirtas system is 800pph and Other system is 500-800pph. In addition, if the 300 dpi scanning is done in color, then the rates are Kirtas 800pph and Other is only 150-400pph!

Variable	Kirtas	Others
Best Case ¹ Rate	100%	100%
Resolution to 300dpi rate	100%	50-80%
36 bit color rate	100%	30-50%
Grays scale rate	100%	40-60%
8.5" to 14" page size rate	100%	70-80%

¹Best Case = black & white, 200dpi, 8.5" wide

The cumulative effect of these variables may cause a significant change in quoted scanning speeds from 100 to 20. This variation must be accounted for in the project cost proposal, or one will have errors in forecasting resource needs, completion time and overall costs!

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Page Capture

The system starts with a digital photograph of the page, using a 16.6 megapixel camera. The camera's optical path is designed to look at one page at a time in the 110 degree cradle. The positioning puts the camera at a 90 degree angle to the page. This position minimizes the curvature effect that occurs near the spine. The text on the page is nearly flat in this position, and the depth of field is nominal so sharpness is maximized.

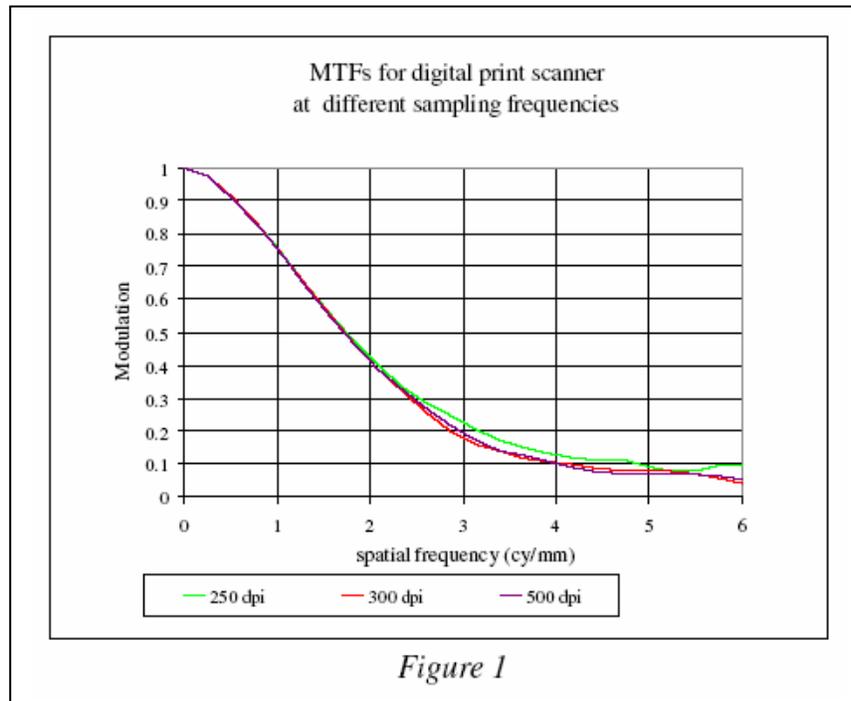
With a digital camera, the digital array is less likely to have imaging defects because the scanner head is stationary and all pixels are sensitized at the same moment. The better the original capture, the better the final image. The camera shutter speed is fast so the page is captured in less than a second compared to 3-6 seconds for linear scanners. This allows the system to be many times faster than others! Real productivity is achievable. And all this is done with full color imagery. All the graphic and picture content in books is faithfully captured for viewing and printing. The APT BookScan 2400 system uses two cameras; one looks at the left-hand pages and the other the right hand pages. Thus the distortion that can occur from page curl near the binding becomes a non-issue even when text is very close to the binding edge. Each camera is connected to a PC via Firewire so the capture speed remains continuous without slowdown or buffering required.

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Resolution

Resolution is an often-quoted measurement associated with image quality. Most people believe that "more is better". Resolution is good up to a point, and beyond that, higher resolution does not continue to add much value or information. While this may seem surprising, the scientific analysis done by Don Williams of Eastman Kodak is quite clear¹. The measure of quality has little to do with DPI. It is more to do MTF (Modulation Transfer Function) and image processing. Resolution is important to a point, and beyond that point, it appears to add negligible value.



*Notice that the MTFs for each sampling frequency (250, 300, and 500 dpi) are essentially identical. The individual curves in Fig. 1 are difficult to identify because they literally overlay. This indicates no real resolution advantage at 300 and 500 dpi compared to the 250 dpi scan. This is indisputable. The 0.1 modulation level corresponds to 4 cycles/mm. Translating this to an effective resolution ($\text{dpi} = (\text{cycles/mm}) * 50.8 \approx 200 \text{ dpi}$), one finds that this scanner is really no better than a 200 dpi scanner no matter what the advertising claims or sampling frequency. This analysis was performed with tools provided through the TC42/WG18 standards group and is one of many examples where they have been used to objectively clarify resolution performance.¹*

The better measurement is the performance of the Modulation Transfer Function (MTF) of the imaging system. This reference provides useful information on the performance of the camera. We use the Canon EOS-1Ds Mark II camera with a 24-70mm zoom lens. The additional reference on the MTF of the digital camera used is also revealing.

http://www.wlcastleman.com/equip/reviews/film_ccd/

¹Debunking of Specsmanship: Progress on ISO/TC42 Standards for Digital Capture Imaging Performance
Don Williams, Eastman Kodak Company, Rochester, New York

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Conclusions

The measurement of the MTF is at 50%, and here are the conclusions.

- * *Current Canon EOS-1D and 20D digital SLR cameras outperform film in image quality and sharpness*
- * *The EOS-1Ds Mark II and the 20D have the smallest pixel size of the cameras tested and have the highest resolution. The EOS-1Ds Mark II achieved the highest image quality when resolution and 50% MTF and image appearance are compared.*
- * *Film continues to have higher resolving capacity with high quality lenses than CMOS sensor-based digital cameras. However, after scanning for viewing and printing, lower 50% MTF in film is associated with lower overall image sharpness.*

The images from the Kirtas systems continue to be the best quality observed.

Image Quality

Nearly all other scanning solutions have a capture system that uses CCD technology for acquiring the image. These systems have either the paper in motion past the CCD, or the CCD in motion passing over the paper. Over time, the image quality will degrade due to the motion factor. Variations in speed and/or vibrations during acquisition will degrade the image quality.

With the Kirtas design, the capture process is under system control improving the results due to the following:

- Image capture is 'still photography' at less than 1/80th of a second. The page is still and flat when the camera captures the page. Nothing is moving. This eliminates image blurriness, and curvature at the spine.
- There is minimal operator intervention during the capture process thus eliminating the presence of hands, arms, etc. in the image.
- The limited operator intervention reduces fatigue that causes other image quality defects.
- Sharper images are generated yielding significantly better OCR processing and text search-able images.

Image Processing

When you start the processing with a high quality image, the processing for deskew, despeckle, rotation, etc., naturally yields the best result possible. During the image processing, the original source image is retained in case some reprocessing or selective processing is desired, thus eliminating the need to re-scan. Image processing can be done in a single book batch mode, or multiple book batch mode overnight.

Image Processing Features of BSE (some, not all)

- Crop image based on mask
- Remove clamp image
- Deskew on text
- Image rotation
- Brightness to lighten or darken
- Contrast to bring out detail
- Image sharpening adjustments to create crisp edges on characters to improve OCR processing results
- Bitonal (DLT algorithm) and grayscale conversion derived from the full color image

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Image Processing (con't)

- Image segmentation to allow color retention of pictures and graphics, while text is in black and white
- Centering of text to the page frame
- Page padding to align the image with the original page size
- Checksum to allow for future checking to detect page alteration
- Page number contact sheet to facilitate one of the QC steps to assure all images are present

All processed images result in a copy from the original raw image file that is retained for subsequent processing or re-evaluations and negating the need for rescans!

Image Segmentation

One of the key features is image segmentation where the processed full color image is segmented into graphics/pictures and text. The graphic/picture areas are retained in full color (or grayscale) and the text is processed to black and white resulting in the optimized image size without loss of content.

Output Format

JPEG or TIFF or PDF outputs are the image format choices to be fully compatible with the image repository systems available.

Metadata (see examples)

The Kirtas system provides Technical, Structural and Descriptive metadata for users wanting this information in *open schema* XML files; data values are preceded by the usual element names (tags).

Technical Data: information about the machine and operation during capture and is automatically generated by the system

Structural Data: information about the book structure/layout requires some manual entry: Table of contents, chapter marks, etc.

Descriptive Data: information about the book is the MARC information that is Dublin Core+2 compatible

The MARC data can be entered with Barcode reading of the ISBN or a user interface for entering content descriptions (e.g., Title, Author, Copyright date, Table of Contents, etc.) within BSE for fast entry and ease of use.

XML Output Format

Metadata is output in a separate XML format file for use with repository databases for potential use in downstream databases or other processing operations.

OCR Processing

Kirtas uses the ABBYY OCR software that is an add-on integrated module. The OCR processing is done after the image processing is completed to achieve the highest yield possible. The OCR output is used to create a text file for indexing, and to be used to produce image-on-text or text-on-image searchable PDF files.

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Image Processing (con't)

Multiple output options

- Choose the interpolated final resolution between 300-600 DPI
- Choose one or more image format output completed in one session as TIFF, JPEG and compressed PDF (2x to 100x)

Book Quality and Image Processing Quality Control:

- The Book Scan Editor software has many features that contribute to excellent image and book quality:
- The process starts with a color image to allow processing with all the information.
- Regardless of features used, one can always return to the original image for re-processing; re-scanning is very, very rare.
- Automated cropping is done separately on left and right pages to eliminate content loss.
- Operator has thumbnail images readily available to check processed image quality; easily and quickly correct as needed.
- The image processing software segments text into black and white imagery while retain color for graphics and pictures.
- Metadata is also available to inclusion with the finished electronic book, including 16 fields of Dublin Core +2 in XML format.

Book Security and Preservation

With the Kirtas APT system, you can take the solution to the books at the library and do the scanning on-site. This means one can scan any rare book in the collection with great care and under security without having to give up possession of the book. The scanning is done with your own people supervising the scanning personnel giving them comfort and confidence in the book handling. Thus, book security is achievable to the level you want to implement.

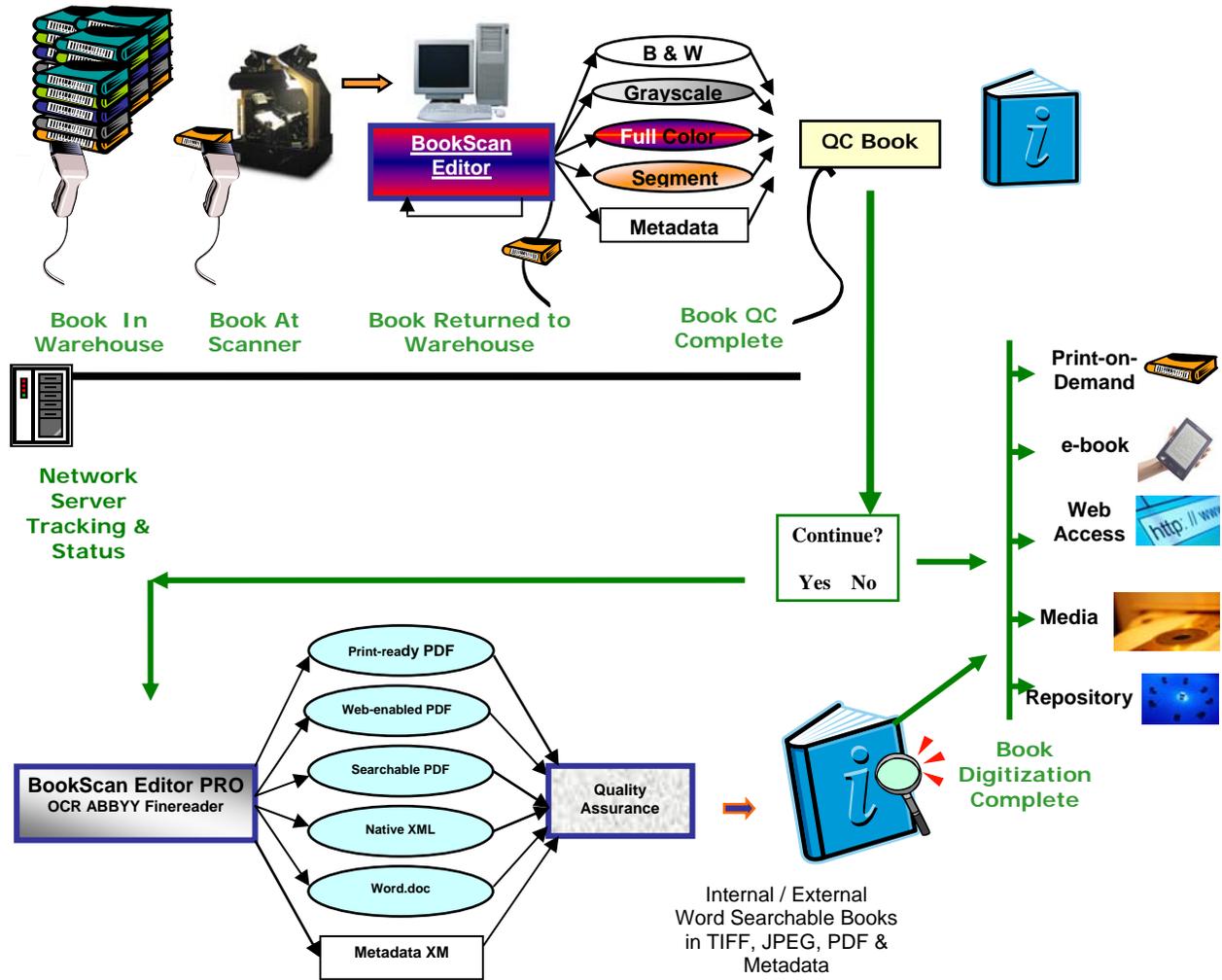
We have tested and scanned acid paper book pages. The robot is gentler than the human hand and less damaging. Kirtas is successfully scanning these types of books for its customers.

The light bulbs are Phillips T41 compact fluorescent and they have negligible UV.

The APT system can be operated in manual mode for books with loose pages or torn pages. The manual mode does not degrade the productivity rate too much.

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.



Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Metadata

An important element of successful digitization is the accompanying metadata about the book that defines its Descriptive, Technical and Structural attributes that are essential for allowing data base cataloguing, maintaining file integrity, page viewing and file activity in creation for audit purposes. For more information on Metadata, use this link.
<http://www.loc.gov/standards/mets/>

The Kirtas digitization process produces metadata to support the needs of the file and for external requirements for creating a digital library. Examples follow.

METADATA NAME	METADATA VALUE
1. Title	Cake Decorating Title
2. Creator	"Sample Creator"
3. Subject	"Sample Subject"
4. Description	"Sample Description"
5. Publisher	"Sample Publisher"
6. Contributor	"Sample contributor"
7. Date	"Sample Date"
8. Type	"Sample Type"
9. Format	"Sample format"
10. Identifier	junk identifier xxxx
11. Source	"Sample Source"
12. Language	"Sample Language"
13. Relation	"Sample Relation"
14. Coverage	"Sample Coverage"
15. Rights	"Sample Rights"
16. Comment	"Sample Comment"

IMAGEPATH - select box to copy current path on File Management page: C:\Project1\ABC1234\9-999-99999-9\0001\001

CURRENTDATE - select box to save: Date:2005-08-19 Time:20:00:18

Descriptive Metadata

Default: Dublin Core + 2 (Image Path and Date)

16 fields fully customizable, typically includes ISBN or MARC data elements. If you are scanning books that are already catalogued that are in a network accessible database, then you can use a barcode reader at the start of scanning to link the book with the existing data elements to preclude key entry and key entry errors.

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Technical Metadata

Generated automatically during the processing and QC steps, and includes all the processing filters applied, as well as image properties and md5 checksum. Please see attached XML sample.

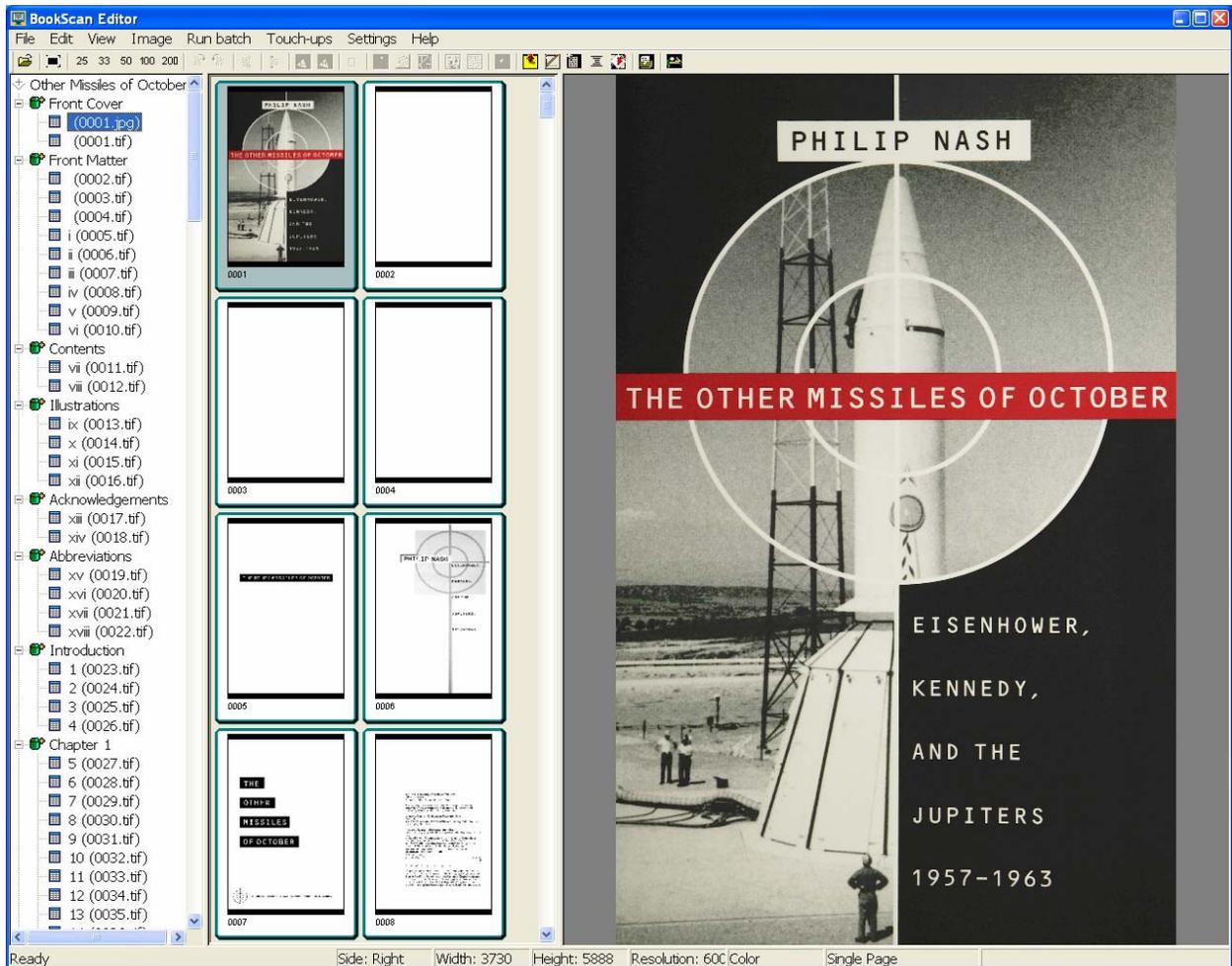
```
<?xml version="1.0" ?>
= <BookInfoDataSet>
= <BookStructTable>
  <ID>1</ID>
  <ParentID>0</ParentID>
  <Name>Other Missiles of October</Name>
</BookStructTable>
= <BookStructTable>
  <ID>2</ID>
  <ParentID>1</ParentID>
  <Name>Front Cover</Name>
</BookStructTable>
= <BookStructTable>
  <ID>3</ID>
  <ParentID>1</ParentID>
  <Name>Front Matter</Name>
</BookStructTable>
= <BookStructTable>
  <ID>4</ID>
  <ParentID>1</ParentID>
  <Name>Contents</Name>
</BookStructTable>
= <BookStructTable>
  <ID>5</ID>
  <ParentID>1</ParentID>
  <Name>Acknowledgements</Name>
</BookStructTable>
= <BookStructTable>
  <ID>6</ID>
  <ParentID>1</ParentID>
  <Name>Abbreviations</Name>
</BookStructTable>
= <BookStructTable>
  <ID>7</ID>
  <ParentID>1</ParentID>
  <Name>Introduction</Name>
</BookStructTable>
= <BookStructTable>
  <ID>8</ID>
  <ParentID>1</ParentID>
```

Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Structural Metadata

Structural Metadata is generated within the BookScan Editor software using the BookView interface. See below a screen shot of a sample book, as well as the corresponding XML output attached.



Bound Book Digitization System and Processing

Why Kirtas offers the best solution in the market today.

Summary

Kirtas is the leader in the market and continues to advance its products, features and productivity to satisfy customer needs with radical, innovative technology.

The Kirtas products and process for book digitization offer you the safest and fastest way to produce high quality digital files of books in a non-destructive manner. The process allows you to create the digital book and the necessary metadata about the book to assure easy import and accessibility within a digital book information management system.

You can be sure that the images produced will always reflect positively on the owner. Sometimes you will only have one opportunity to digitize a book, and Kirtas will assure you are successful.

Kirtas - Moving Knowledge from Books to Bytes