

Project Report: CARIBBEAN NEWSPAPER IMAGING PROJECT

Phase I: Imaging and Indexing Model

By Erich Kesse, Robert Harrell, Richard Phillips, and Cecilia Botero

Abstract

This paper describes the University of Florida's [Andrew W. Mellon Foundation](#) funded *Caribbean Newspaper Imaging Project*: its goals, approaches and achievements. The Project, designed to convert newspaper microfilm holdings to electronic images, is described in context with previous preservation effort, together with discussion of the limitations of microfilm as an access technology. A review of progress toward goals establishes Project strategies while modeling the implementation of electronic imaging guidelines and the adaptation of traditional technical skills from both cataloging and analog imaging. Critique, particularly of pitfalls and failures, suggests areas for future consideration.

Understanding the Past

Florida, its influx of immigrants from and volume of trade with the Caribbean, is almost as much a member state of the Caribbean community as of the United States of America. In Florida's research libraries, emphasis on the collection and preservation of Caribbean resources has a long history rivaling that of Floridiana. The University of Florida, in particular, maintains a large and rich collection of Caribbean archives and publications. The collections are important to building an understanding of the region, bridging cultures, and fostering economic ties. More recently, the collections, by virtue of their preservation in microfilm and the loss of source-documents, have come to represent extensions of various national archives. Legislative reports published by the government of Guyane Française and microfilmed by the University of Florida, for example, continue to exist only in microfilm.

The University of Florida began collecting Latin American and, particularly, Caribbean research resources in the late 1920's. U.S. interest in the region at the time, already attenuated by administration of Cuba at the end of the last century, had been heightened by its occupation of Haiti beginning five years earlier in 1915. Following World War II and the convergence of the Farmington Plan¹ with the application of microfilm technology, a dedicated faculty and staff systematically built a vast collection - today, more than 1.5 million items - of Caribbean government documents, journals, manuscripts and archives, maps, monographs, and newspapers. In its Latin American Collection alone, the University of Florida holds more than 300,000 volumes of printed materials; a growing number of electronic resources; nearly 50,000 reels of positive microfilm; and, in preservation storage, more than 8,500 reels of negative microfilm masters. The latter represent more than 5 million exposures or 9.5 million pages. The fact that 7,000 reels of microfilm masters are newspaper holdings indicates the collection development and preservation effort emphasis.

Newspaper microfilming began in earnest in 1953. The Rockefeller Foundation funded a technician, traveling throughout the Caribbean, with a portable microfilm camera to film materials that could not be acquired otherwise. Many of these materials, today, continue to exist only in microfilm. The University of Florida's microfilm masters are the archive of several newspapers among them Cuba's *Diario de la Marina* and Haiti's *Le Nouvelliste*.

By the 1960s, supported by state funds, fed by standing orders, and empowered by copyright legislation known as the Inter-American Agreement (1939), a program of microfilming Caribbean and Florida newspapers had been established. Today, the program, which operates under national guidelines and standards for production, duplication and archiving of microfilm for preservation, continues albeit more restricted by changes in international and U.S. copyright legislation. Long standing agreements between the University of Florida and University Microfilms International ensure the availability and continued preservation of these materials as originally envisioned by the Rockefeller Foundation and the Farmington Plan.

The Problem

Microfilm technology advanced the collection and distribution of resources. Today, it remains a reliable and cost effective means of long term preservation. Microfilm continues to be the medium of choice for stability, life expectancy and image quality and especially for large-format, small-font or fine-line source-documents such as maps and newspapers. Microfilm's several limitations, however, afford it the distinction of least respected information delivery format². Microfilm must be used *in situ* and, usually, without the benefits of indexing or relatively immediate image retrieval afforded by newer automated information delivery formats.

Perhaps most limiting, microfilm is difficult to maintain and expensive to replace. Microfilm deterioration begins whenever optimal environmental conditions or microfilm readers are not adequately maintained. Attaining optimal conditions, particularly difficult in Florida and the Caribbean basin countries that rely upon the microfilm, incurs its own high cost;³ the heating, ventilating and air conditioning (HVAC) control systems required are neither inexpensive nor easily maintained. Increasingly, as well, the cost of maintaining readers to service the microfilm is becoming difficult to bear. Once ubiquitous microfilm readers and reader-printers are losing market share to multipurpose and more ubiquitous computers. Replacement parts and service personnel for microfilm readers/reader-printers are increasingly few. Taken together, the costs of acquiring, maintaining, servicing and replacing microfilm is becoming prohibitive particularly throughout the Caribbean where poor climate and weak economies converge.

The challenge, which the University of Florida and the Andrew W. Mellon Foundation seek to manage through the Caribbean Newspaper Imaging Project, is the development of an electronic global resource sharing model, both feasible and economical, for information in newspapers. Born of ideas defined by Yale University's Project Open Book⁴ and the University of Michigan's now independent Journal Storage Project

(JSTOR)⁵, the Caribbean Newspaper Imaging Project is at once hybrid and new. Stated Project goals⁶ are these:

1. Convert approximately 132,500 microfilm exposures, the record of two newspapers: Cuba's *Diario de la Marina* and Haiti's *Le Nouvelliste*,⁷ to digital images;
2. Provide multi-lingual indexing in the newspaper's native language (i.e., Spanish and French) and English;
3. Implement cost recovery marketing in order to support conversion of additional titles; and
4. Establish efficient, low cost models for facilities and productivity; which would allow other institutions to share the burden of newspaper digitization.

Project completion would require examination of several additional issues.

Microfilm to Digital Conversion Issues

Conversion issues were several: selection and configuration of a facility; file characteristics and directory structures; source-document definition and condition, and work-force issues among them.

Archives and Distribution

The definition of an archive was primary. As in Project Open Book, the microfilm would remain *the* archive of source-document; both qualities of images and life expectancy under optimal storage conditions were known.⁸ Multiple master storage sites and a monitoring program based on national standard⁹ would ensure continued preservation. Moreover, the resolution of digitized images of newspapers would only approximate that of the microfilm.¹⁰

To safeguard investment in the digital product, DAT (i.e., digital tape) would archive the electronic files with an additional copy maintained in CD-ROM, the format elected for distribution. Electronic archives would be placed in storage conditions meeting existing standard and monitored in accord with existing industry standards. In many ways, the management of an electronic archive has been with us for more than a decade in the form of locally held automated catalog-record tapes, census information, and other electronic files.

Distribution of images via the Internet was considered but rejected during the planning process. Internet distribution for both project titles would have required in excess of 197 GB of active storage space not available at the Project's start. Moreover, conveyance of the images had its own problems. Though GIF-on-the-fly software would have made images browsable without additional labor, GIFs were large enough, in terms of bytes-per-image, to render remote access laboriously slow without large and dedicated bandwidth. The graphical size of images was yet another problem. Images would not fit,

legibly, within a browser's viewing pane; awkward bi-directional scrolling was required. Further, GIF's "lossy" conveyance reduced image quality. This was evident, particularly, in image areas most dependent on fine resolution such as the classifieds. While we continue to investigate Internet distribution, it was and remains our conclusion that this form of distribution will not be viable until the problems listed above can be resolved. Distribution of TIFF images bundled with a TIFF viewer and an index interface on a CD-ROM, conforming to ISO 9660, **12** was elected.

Facilities

The decision to build an imaging facility within the University Libraries was made during the planning stage. At the time, the number of commercial facilities offering microfilm conversion services was few and the fees charged by existing services was not considered to be economical. The University's Preservation Department had the requisite managerial and production experience, with its in-house microfilming facility,**13** and had been building the networking experience necessary to establish an in-house digitizing facility. Additional knowledge of electronic imaging and digital formats was gained through the Cornell University Digital Imaging Workshop,**14** together with an exhaustive program of reading and experimentation. Characteristics of the space needed were similar to that housing the Department's microphotography facility. A vibrationless, dust-free environment, darkened independently of adjacent offices was carved from existing space.

Microfilm scanning equipment selected by the University of Florida would have to support intensive long-term use and produce images meeting a high image quality threshold as suggested by Project Open Book and the Cornell Workshop. Equipment also would have to be affordable in terms of producing images at the lowest possible cost. Several microfilm scanners capable of meeting the quality requirements were available but would have increased the final per-image cost several fold. The Mekel scanner, with software components, used by Project Open Book, cost more than \$100,000. The Minolta MS1000 scanner, including software, with which the Caribbean Newspaper Imaging Project was begun, cost less than \$25,000. A second scanner, the Minolta MS3000, was added to meet production targets less than one year after purchase of the MS1000 at less than \$21,000.

The Minolta products provided acceptable dots-per-inch (dpi) resolution and gray scale. They lacked the Mekel scanner's several automated features, but these were deemed unnecessary owing to characteristics of the selected newspaper microfilms. The Minolta equipment was capable of scanning to a depth of 400 dots per inch (dpi), regardless of filming mode, but depended on resolution of the image projected on screen at the time of imaging. The Mekel equipment, in comparison, was capable of scanning materials filmed in two-up comic mode at 300 dpi and those filmed in two-up cine mode at 600 dpi.**15** It had no dependence on projected screen resolution; images were made directly from the film. Characteristics of the microfilm (i.e., two-up comic mode) muted questions of selection. The Minolta equipment was sufficient if not, in some ways, more versatile for scanning newspapers on microfilm in two-up comic mode.

When the project began, a 486 CPU, 66 MHz workstation was the best available computer to drive the scanners. Each workstation ran with 8 MB RAM and temporarily saved scanned images to 2 GB hard-drives. While this configuration was adequate for Project start-up, it was quickly determined that a more powerful configuration was needed to increase productivity over scan-time. Each of the scanner workstations has been up-graded to Intel Pentium CPU, 166 MHz, running with 32 MB RAM. Workstations were also outfitted with 20-inch monochrome monitors to facilitate image quality assessment. In addition, uninterrupted power supplies (UPS) became standard for all scanners and back-up workstations, as well as for the server, guarding against electrical malfunction, lightning strike, etc.

Working under a distributed computing model, other equipment was selected for remote indexing; 4mm DAT backup; and CD-ROM distribution-product creation. Microfilm scanners and other equipment were added to the Preservation Department's existing local area network (LAN), an Intel Pentium CPU, 166 MHz server with 128 MB RAM, running NOVELL 3.11. A subsequent hardware up-grade and migration to a Windows NT platform increased speed and file management capabilities. At the project's start, the LAN consisted of 10 Windows 3.11 and Windows 95 workstations, connected by thin-wire Ethernet, since up-graded to a dedicated hub using twisted pair, fast Ethernet. The server has an 8 GB storage capacity with 4 GB dedicated to image file transfer, assessment, etc. This capacity is sufficient for file processing only and requires nearly constant file archiving. Throughput needs demanded similar attention be paid to bandwidth. Bandwidth limitations necessitated transfer of images from server to the remote mastering workstation equipped for both CD-ROM mastering and DAT backup.

Because of the magnitude of the files and the complexities of maintaining multiple user access for inputting index records, in house digitizing requires a significant commitment of Systems staff. Networking and workstation requirements should be given serious consideration even for those programs that opt to out-source scanning. The facilities and physical support structure required to perform image quality review alone is not insignificant.

File Characteristics

File characteristics include scan depth; tonal qualities; file format; and compression. For optimal image quality in library applications, these characteristics are defined by the emerging standard established by the Cornell University Libraries.**16** Scan depth (i.e., dpi) and tonal qualities determine resolution.**17** File format and compression determine file size and "lossiness."**18**

It was determined that source-documents would be imaged at 400 dpi with 64 levels of gray, the maximum level allowed by the Minolta scanners.**19** The microfilm used for newspaper filming is a high contrast medium which is essentially bitonal. Use of gray scale in imaging would maintain any tonal qualities captured by the film in illustration and fine or small print.**20** Scanned files would be saved in the tagged image file format (TIFF), using ITU T.6 (formerly, CCITT Group 4) compression. TIFF images with ITU

T.6 compression are "lossless." File sizes, ranging between 0.8 and 1.4 MB compressed, and the number of files to be saved, more than two hundred and sixty five thousand, obviated saving files uncompressed. With compression, there was a nearly one-to-one conversion. Production generated, on average, approximately one CD-ROM for every reel of microfilm converted.

Article data tables used for indexing and abstracting were built as a FoxPro relational database application. Delphi programming was used to build both a multi-user interface for access to index and abstract entries and a viewer for access to images. Data elements allowed record of newspaper and article titles; enumeration, pagination and column numbers; author; subjects/index terms; and publication chronology and event dates, as well as, searchable keyword abstracts in English and the newspaper's native language, French or Spanish.

Directory Structure

Newspapers are readily adaptable to a directory structure that is intuitive to any user insofar as their chronology suggests structure. Directories are arranged with title at the top level, followed in cascading order by year of publication, month of publication, date of publication, and section-and-page number. The front page of the *Diario de la Marina's* June 1, 1956 issue, for example, equates to the file located at **[drive letter]:/Nouvelliste/1956/06/01/A01.tif**. This scheme works well, in turn, when querying or parsing requests from index-interface (i.e., relational database) and image-viewer programs.

This scheme, however, does not easily accommodate page-name anomalies in the source-document. Failure to anticipate anomalies aside, **21** anomalies that occur as a result of printing or publication can be "corrected" only through indexing. Under the distributed computing model employed by the Project, correction through indexing requires coordination among indexing and imaging staff in referencing and naming anomalous files. Misprints resulting in incorrect publication of chronology and pagination require corrective action that is similar to but more proactive than attention shown to correct such problems during microfilming for preservation. Without indexing, directory structure and file naming conventions that do not impose a consecutive image numbering scheme are unforgiving of anomalies in chronology and pagination. At the same time, consecutive image numbering schemes without indexing prohibit intuitive image access; images must be "paged" or viewed image by image. Our experience suggests that a directory structure and file-naming scheme be standardized for serialized information and, particularly, for information in newspapers.

This scheme also is not favorable to the preservation practice of microfilming a single page at multiple densities, one optimized for the capture of text and the other optimized for the capture of illustration. In this Project, exposures optimized for text capture were deemed most important and, therefore, scanned and saved with the standard directory/file-name designation. Illustrations were rarely indexed, though several notable and important illustrations were recorded. If the microfilm does indeed capture graphic

information better than the digital version, conversion of the exposure optimized for illustration might not serve its intended purpose. When the exposure optimized for illustration was scanned and saved, it was saved with additional designation, e.g., **A01a.tif**. Because files saved with the additional designation could not be parsed by the index-interface or viewer programs, their value was almost solely for purposes of quantifying differences between the microfilm and digital versions. No thought was given, beyond an initial test, to pasting the scan of the optimized illustration into the scan of the optimized text; the size of the relative parts was greater than the resources of the individual workstations (i.e., their CPU, RAM and virtual RAM).

It would also be advantageous to standardize, beyond the experience of this Project, the data-elements used during indexing and abstracting of newspapers. While the practice of this Project was to record information in a relational database that treated the image file as an object in a table, this information could be recorded as Standardized General Mark-up Language (SGML), metadata, or other file header information. The database method's advantage is that imaging and indexing can proceed separate from or in advance of imaging, assuming agreed upon methods of relating the image object to the index. It afforded time to review entries by area and language specialists working at their own pace. Image objects could be committed to the electronic archive immediately. Other methods build an index through tagging an existing image. This would have required either indexing as imaging occurred or, more aptly suited to the Project's distributed computing model, maintaining images in active disk space or an intermediary file until tagged. Immediate indexing would have necessitated either in-put ready indexing or a staff with an unlikely combination of imaging, indexing and language skills. Preparation of in-put ready indexing would have required additional start-up time, which was not available. Maintaining images in active disk space would have required additional server and bandwidth resources, which would have slowed progress and decreased cost-efficiency. Use of intermediary files would have necessitated an additional layer of tracking and management.

A software interface, programmed specifically for the project, allows users to search the index and abstracts and to browse images. The directory's structure is used to link index and abstract entries with the image objects. While this software, particularly the freeware TIFF image browser, was necessary to complete the project, it likely, soon, will become a once convenient but no longer necessary tool of the past. Standardization of newspaper article indexing and abstracting data-elements and the subsequent mapping of these elements as a SGML or XML Document Type Definition (DTD), perhaps with crosswalks to other DTDs, will make the software obsolete.

Source-document Issues²²

The source-document issues resulting from filming were several. Issues related to the source-documents and the source-document microfilms had to be considered. Printer's effects; shipping, binding and storage effects; embrittlement effects; paper characteristics; and illustration and font sizes were issues of concern regarding the source-document. Source-document lighting; processing and storage effects;

orientation; reduction; exposure and density; and resolution were issues of concern regarding the microfilm. Planning for the work required assessment of the source very much as would have been necessary to microfilm a source or generate paper facsimile from a microfilm. Traditional use of random survey and interpolated data was made during planning. In retrospect, more detailed analysis was required. The great variety of source-document and microfilming characteristics proved assumptions based on survey to be inaccurate. The sample's $\pm 10\%$ level of confidence was inadequate.

The adage, "garbage in, garbage out," is a harsh solipsism to say that electronic technologies cannot reverse defects borne onto microfilm. Scanned from source-documents, image defects such as staining or those resulting from creases and folds are discernable from the text they obscure by gradient differentiation techniques.

Once committed to a high contrast medium such as microfilm, however, differentiation between defect and text becomes unlikely. Text readable through stains on the original are often no longer readable on microfilm.

Effects such as bleed-through, transference, and uneven or over-inking had to be noted in order to assess the quality of individual scans. Minor but time consuming corrections to improve hardware and software performance had to be made throughout the Project. The nature and number of corrections demonstrated uniform conversion settings to be arbitrary and would have rendered automated features of the Mekel equipment useless. More detailed assessment might not have reduced this burden but would have assured Project managers both of initially adequate staffing funds and a workforce, trained, from the start, to deal with the broadest range of image defects.

The titles selected for conversion were microfilmed between 1957 and 1987. Some were microfilmed by the University of Florida in their country of origin on portable equipment and others, at the University of Florida on stationary equipment. While the most consistently reported physical defect encountered was scratching, deterioration of the microfilms' acetate base was evidenced by tears, curling and separation of the emulsion from the base throughout the microfilm collection. Every imaginable effect of filming practices also was encountered. The thirty years between 1957 and 1987 was a period of increasing standardization; both the growth toward standard practice and every change in standards can be seen on the microfilms, together with the defects of filming. Even defects such as slight light imbalance on the surface of the source-document during filming become troublesome during scanning of newspapers reduced twenty-one times onto microfilm.

Not all problems noted could be corrected. Image enhancement techniques, e.g., dithering, despeckling, etc, could not be used effectively owing to the nature of high contrast microfilm or the fine resolution of broadsheet newspapers on 35 mm microfilm. Removal of scratches and errant marks, for example, could not be automated without the loss or degradation of text. Manual removal was not cost effective. Moreover, when manual correction was completed, the task often required native language skills. In review, the exercise proved pointless; native language readers were able to adequately

discern words from obscured text. Though enhancement and human intervention will likely remain a necessity if intelligent character recognition (ICR) or optical character recognition (OCR)²³ are to be employed on scanned newspapers, improvements in software first must make the task more efficient and cost effective. Corrections, which were cost effective, were largely mechanical and similar to those undertaken during microfilming. Alignment problems, for example, required rotation or deskewing. Source-document microfilm density problems, including over and under exposure as well as inking effects, could be minimized by manipulating lighting conditions during scanning.

Some problems were the result of the mechanism. Residue of spent filaments inside the vacuum of the scanner's bulb produced image effects, for example, which required the workforce to build expertise, differentiating the effects of bulb condition from unbalanced inking, wearing or exposure.

Quality control of scanned images was performed through a process of benchmarking, which made visual comparisons between the digital quality of an optimized image and successive images, a method similar to that developed by Yale.²⁴ Differences in method were necessitated by differences in microfilms and source-documents. Project Open Book assumed that scanned microfilm met or closely approximated current standard and contained images of average book size filmed at reduction normal for books. The Caribbean Newspaper Imaging Project microfilm was produced prior to current standard and contained images reduced more than twice that required for book microfilming.

The unit against which image quality comparisons were made was the smallest "e," usually in the classifieds, of the microfilm. Benchmarking required "optimizing"²⁵ the page containing the "e" and comparing the clarity of text on scans subsequent to it. Benchmarking, as "quality e measurement" in microfilming for preservation, was done approximately every tenth image. Images were optimized approximately every 300 scans or as the work-force changed. Benchmarking was partly art, requiring subjective judgment, particularly when image density varied across a single page. Different scan settings only improved the legibility of different parts of a given page.²⁶ As much of the microfilm that libraries depend upon has not been produced to the level of current standard, problems associated with conversion of substandard microfilm require further consideration.

Display size of the scanned source-documents represented additional problems. Display at a one-to-one ratio was too large to fit and easily navigate on screen. Display at reduction to fit or navigate easily on screen rendered text illegible. The solution was programming of a TIFF viewer containing a "magnifying glass".²⁷ Images are opened to fit on screen in a "window" containing a magnifying glass that can be moved by dragging the device over the image. The image area beneath the magnifying glass is displayed legibly in a separate window. This solution also resolved the problem of fees and legal agreements associated with embedding image viewer software on the CD-ROM with the images and index; the use of a viewer programmed by the Project would incur no additional costs.

Workforce Issues

Other than its indexing component and the use of older microfilms, the Caribbean Newspaper Imaging Project most differed from Yale's Project Open Book in staffing.²⁸ Trained and managed by permanent staff, student assistants were hired to perform the bulk of tasks. Students were available in a large pool, inexpensive, easily trained and often highly computer literate or fluent in French or Spanish. While use of a student workforce had its disadvantages, e.g., high turnover, high levels of supervision, retraining, scheduling, consistency of product, its pay-off was in low cost. Student staffing reduced costs to near two-thirds that which permanent staff would have incurred. Intensive training and review of performance and products assured quality and consistency of product while lowering per image costs from those calculated for the employment of full-time staff during project planning.

Indexing routines were supervised and work reviewed by three Latin American and language specialists who also defined indexing criteria and the select, controlled vocabulary derived from *Library of Congress Subject Headings*. Approximately 2 FTE part-time staff was employed to index and abstract. Part-time staff was paid \$6.50 per hour and did not accrue benefits. Native French speakers, mostly from Haiti but also from the French Caribbean and French north and west Africa, indexed and abstracted articles from *Le Nouvelliste*. A small pool of available French speakers slowed completion of the task. Native Spanish speakers, largely of Cuban descent, indexed and abstracted the *Diario de la Marina*. In both cases, indexing and abstracting were done in the native language and later translated into English, completing bilingual indexing requirements. More than 20,000 articles were indexed and a minimum of one article per issue was abstracted. Article selection was at the discretion of the indexer/abstracter within criteria established by the Latin American specialists. Quality control and editing were subsequently completed by the specialists.

Imaging routines were established and images reviewed by a reprographics specialist who also managed DAT archiving and CD-ROM production. Approximately 2 FTE part-time staff, a sufficiently stable workforce, was employed to image the microfilm. Part-time staff was paid \$5.00 per hour (i.e., slightly above the minimum wage at that time) and, for the most part, did not accrue benefits. Two-hour shifts were maintained in order to optimize attention and minimize the risks of eye-strain and repetitive stress syndrome. The average employee scanned at a rate of 1.25 images per minute (IPM). Those staff whose productivity was low - 0.5 IPM was the lowest recorded - or whose accuracy or image quality were consistently low were dismissed. The reprographics specialist, who worked regular shifts to maintain skills and demonstrate efficiencies, was frequently able to produce images of acceptable quality at rates in excess of 2.75 IPM. Most efficiencies, other than those gained through networking up-grades, were achieved through mechanical means, e.g., film advance techniques. Other measures such as the two-hour shift, however, resulted in equal gain. With both microfilm scanners operating, an average of 1.5 GB of scanned images was produced each day of operation. Scanners operated between 65 and 120 hours per week.

Systems support staff included FoxPro and Delphi programmers, as well as, a network trouble-shooter. Attempts to hire a computer programmer to develop both the multi-user indexing system and public-user interface were fruitless. State of Florida staffing plans had been unable to compete with corporate market forces, leaving Systems Department programmers to assume responsibility at the cost of delay in other project schedules. Network software was configured by Systems staff but administered by Preservation staff; the network actually pre-dated the Project and was expanded to accommodate it. Insofar as programmers' work may be borrowed or adapted, other projects working from the experience of this Project should not require as much or the same type of programming assistance. Networking requirements, hardware, and bandwidth use grew rapidly throughout the Project and were associated predominantly with up-grades to increase performance. Networking speed was the single most important factor in increasing productivity and decreasing costs.

Project Statistics and Costs

Caribbean Newspaper Imaging Project digitization of *Le Nouvelliste* and *Diario de la Marina* comprises more than 20,000 index entries, 40,000 abstracts and 265,000 images. In total, indices, abstracts and images occupy more than 200 GB. Images, alone, fill 98 archived 2 GB DAT or 329 distribution-ready 650 MB CD-ROMs. CD-ROMs contain images, a viewer, and indices and abstracts for the images on each CD-ROM. Images are available by title, date or subject, supplied on CD-ROM, with other distribution formats negotiable.

Project costs were calculated to include labor, media and equipment costs. Labor costs included wages, salaries and benefits paid to part-time and full-time staff for indexing and abstracting, imaging and related functions, and software development and network support. The table, below, is a summary accounting of expenditures per image.

EXPENDITURE CLASS	COST Per Image
Media (DAT and CD-ROM)	\$ 0.01
Hardware & Software	\$ 0.11*
Scanning & Archive Mastering	\$ 0.08
Indexing & Abstracting	\$ 0.08**
Programming & Systems Support	\$ 0.16
Project Administration	\$ 0.06

Total per Image Cost	\$ 0.50
----------------------	---------

* Hardware and software costs including purchases and up-grades were based on equipment life of five years and prorated for the life of the Project. Of the total hardware and software costs, \$ 0.10 per image supported scanning and archive mastering; \$ 0.01 per image supported indexing and abstracting.

** Calculated per article indexed and abstracted, the actual cost of Indexing & Abstracting was \$0.56.

In relative terms, imaging costs are comparable to those reported by Yale.²⁹ Comparison with Project Open Book is not exact; differences in the type of source documents, the quality of source microfilms, and the selection of equipment to achieve their ends prohibits true comparisons. Caribbean Newspaper Imaging Project cost reports excluded network storage, transaction, maintenance fees and wire costs which might have been included had a network not been previously owned and operated. These costs also appear to have been excluded from summary data produced by Yale.

The Caribbean Newspaper imaging project is a cost recovery project by design both as incentive to efficiency and as a means of expanding the project to subsequent titles. Assessment of efficiencies is still on-going. Problems experienced as the model was implemented, however, suggest its imperfection. Indexing and abstracting, in particular, proved more costly than anticipated. At fifty-six cents per article indexed and abstracted the model demands an alternate approach. Bilingual abstracting, in particular, appears economically unfeasible.

Conclusion

The Caribbean Newspaper Imaging Project establishes yet another model for digitization, one of the first to deal with newspapers on microfilm. Among Project goals, only cost recovery through sales has yet to be achieved. In some ways, the creation of a large image viewer for example, the Project exceeds its goals. The Project, while not directly comparable to other implementation demonstrations such as Yale University's Project Open Book, provides summarized cost data on par with the most cost efficient of those projects.

The Caribbean Newspaper Imaging Project builds new experience for digitization of texts from microfilm predating current "standard" practice. It suggests means of classifying and naming, indexing and abstracting newspapers and places a price on these practices, albeit high. Building on this Project, related secondary projects, such as the on-going Eric Williams/*Trinidad Guardian* Project, explore the possibilities and costs associated with optical character recognition (OCR), adding full-text for select, highly significant articles.

The technical experience of this Project and other projects warily suggests that microfilming guidelines be reviewed and revised for the benefit of future digitization. At the time current standards were written, microfilming was a child we wanted to raise correctly. Today, microfilming has entered an adulthood, about to become a parent whose bad habits may be passed on to the next generation of technology's products. In recent months, at its summer 1997 meeting, the Association of Research Libraries has authorized a task force to investigate this suggestion. It is hoped that the reports of this task force will effect changes in the practice of microfilming which will optimize and further reduce costs associated with digitizing microfilmed source-documents including newspapers.

Endnotes

1. The Farmington Plan was a cooperative collection development plan begun in 1948 and joined voluntarily by American libraries as a means of increasing the number of resources, largely of foreign origin, available to researchers in the United States. The University of Florida assumed "country responsibilities" for materials published in the Caribbean basin. With its presence in the Seminar on Acquisitions of Latin American Library Materials (SALALM) and the Latin American Microfilming Project (LAMP), the University continues to meet these responsibilities.
2. Cf, Anderson, Arthur James. "Faculty to library directory: we hate microfilm." *Library Journal*, v.113 (Oct. 15, 1988), p.50-52.
3. While the University of Florida stores microfilm masters under exacting conditions prescribe by national standards (cf, <http://kamelik.uflib.ufl.edu/repro/micrographics/manuals/storage1.html>), its storage of microfilm for research use is optimized for human comfort and inadequate for microfilm longevity.
4. The Commission on Preservation and Access has published information about Project Open Book. Cf,
 - a. Waters, Donald and Shari Weaver. *The organizational phase of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery*. (Washington, D.C.: Commission on Preservation and Access, 1992). Reprinted in: *Microform Review*. v.22,n.4 (Fall 1993), p. 152-159.
 - b. Conway, Paul and Shari Weaver. *The setup phase of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery*. (Washington, D.C.: Commission on Preservation and Access, 1994). Reprinted in: *Microform Review*. v.23,n.3 (Summer 1994), p.110-119.
5. Cf, the JSTOR web site at <http://www.jstor.com/>
6. Additional information about the Caribbean Newspaper Imaging Project and its goals may be found at the Project's web site, <http://kamelik.uflib.ufl.edu/projects/mellon/>

7. These titles were selected from the more than 100 in the University's archive of newspaper microfilm masters because of their relevance to current events and the importance of their countries of origin in the affairs of the United States and the history of the Caribbean basin. For more information particular to the selection of each title, see the Project's web site.
8. Cf, Lauder, John. "Digitization of microfilm: a Scottish perspective." (*Microform Review*. v.24, n.4 (Fall 1995), p.178-181.)
9. Association for Information and Image Management. *Standard for information and image management : recommended practice for inspection of stored silver-gelatin microforms for evidence of deterioration*. (ANSI/AIIM MS45-1990) Silver Spring, MD : the Association, 1990.
10. White, William. "Image quality in analog and digital microtechniques." (*Microform Review*. v.20, n.1 (Winter 1991), p.30-32.
11. The Minolta Corporation's free TIFF viewer plug-in for Internet Explorer and Netscape
(cf,http://www.minoltausa.com/low/static/tiff_plugin/tiff_view.html) alleviates some of the problems associated with both image size and browser access to TIFF files, but does not reduce download time; TIFF files are larger than those of other file formats.
12. International Standards Organization. *Information processing -- Volume and file structure of CD-ROM for information interchange*. [ISO 9660:1988] Geneva, Switzerland: the Organization, 1988.
13. The Preservation Department produces more than 500,000 exposures annually. Its managerial staff, who have served on industry and library standards committees, oversee the production of microfilm in compliance with American National Standards Institute (ANSI) and Association for Information and Image Management (AIIM) standards and Research Libraries Group guidelines.
14. The Workshop manual, authored by Anne R. Kenney and Stephen Chapman, has been published as *Digital imaging for libraries and archives* (Ithaca, NY: Cornell University Library, 1996).
15. Conway, Paul and Shari Weaver. *The setup phase of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery*. (Washington, D.C.: Commission on Preservation and Access, 1994), p.15. Reprinted in: *Microform Review*. v.23,n.3 (Summer 1994), p.115.
16. Kenney, Anne R and Stephen Chapman. *Digital imaging for libraries and archives*. (Ithaca, NY: Cornell University Library, 1996).
17. *Resolution as it relates to photographic and electronic imaging*. Technical report, TR26-1993. (Silver Spring, MD: Association for Information and Image Management, 1993).
18. For definition, see: *Glossary of imaging technology*. Technical report, TR2-1992. (Silver Spring, MD: Association for Information and Image Management, 1992).

19. Initially, Minolta software allowed a maximum of 16 levels of gray. Though early images from the *Nouvelliste* were made at 16 rather than 64 levels of gray, the difference is minimal, most tonal quality having been lost as a result of microfilm.
20. Many images made from the *Diario* were bi-tonal rather than gray-scale. High contrast microfilming, necessitated for the capture of its faint print, virtually reduced illustrations to black and white. Bi-tonal imaging resulted in savings of file space which out-weighed the slight advantage of gray-scale imaging in this case.
21. Conjunction of section letters with page numbers (e.g., A01, A02) in the file name results as a failure to fully review and define the characteristics of publication. While reasonably intuitive, the conjunction requires additional programming in the index-interface and image-viewer programs to distinguish and correctly query and parse numeric and alpha-numeric file names.
22. For a more detailed description of source-document issues, see: Conway, Paul and Shari Weaver. *The setup phase of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery*. (Washington, D.C.: Commission on Preservation and Access, 1994), p.6-9. Reprinted in: *Microform Review*. v.23,n.3 (Summer 1994), p.111-112.
23. For definition, see: *Glossary of imaging technology*. Technical report, TR2-1992. (Silver Spring, MD: Association for Information and Image Management, 1992).
Application of ICR or OCR on imaged newspapers, especially those converted from microfilm is problematic also for other reasons, principally, the digital resolution requirements of software currently available. The University of Florida is currently modeling an OCR application for newspapers converted from microfilm; results may be seen in its Eric Williams/Trinidad Guardian Reporting Project web site:
<http://karamelik.uflib.ufl.edu/williams/guardian/>
24. Conway, Paul and Shari Weaver. *The setup phase of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery*. (Washington, D.C.: Commission on Preservation and Access, 1994), p.10-11.
25. "Optimization" entailed clarifying the digital image through manipulation of scan-settings. Scans of the image containing the "e" were enlarged, sometimes to the point of pixelation; the scan with the best settings produced the least blocking. Periodically, images were printed out and compared as described by Yale, but this method produced results no better than had been produced by visual comparison of on-screen enlargements.
26. Albeit, as a single setting per frame. Minolta equipment does not support "windowing," i.e., the ability to optimize for illustration with one setting and for text with another setting in one scan. Yale reports similar limitation with Mekele equipment; cf, Conway, Paul and Shari Weaver. *The setup phase*

of Project Open Book: a report to the Commission on Preservation and Access on the status of an effort to convert microfilm to digital imagery. (Washington, D.C.: Commission on Preservation and Access, 1994), p.15. Reprinted in: *Microform Review*. v.23,n.3 (Summer 1994), p.9.

Use of image composition software, e.g., Adobe Photoshop or Paintshop Pro, to achieve this result both was cost prohibitive and yielded inadequate results. The high contrast medium of microfilm had irrevocably damaged tonal qualities of most illustrations.

27. The TIFF viewer was made available only on page-image CDs [no longer available - CNIP contents will be migrated to the Internet in the future]. Its interface has been programmed for the Project, but also allows use with other large digital documents such as maps.
28. Cf, Conway, Paul. "Yale University Library's Project Open Book." *D-Lib magazine* (February 1996) [published electronically at: <http://www.dlib.org/dlib/february96/yale/02conway.html>] for discussion of staffing.
29. *Ibid.* Project Open Book did not incur indexing and abstracting costs as did the Caribbean Newspaper Imaging Project. Caribbean Newspaper Imaging Project cost reporting separates indexing and abstracting costs from imaging costs in order to establish some degree of comparability.