A Cooperative Approach to Document Imaging, Storage and Retrieval
Harold T. George
St. Mary’s University
San Antonio
Texas

Abstract

The immediate objective of the St. Mary’s University Document imaging project is to provide the Office of Financial Assistance with the ability to scan, store, index, retrieve, view and print the images of documents they receive.

The overall objective of this project is to provide this same ability to a number of departments within the University. Those departments would include, Admission, University Relations, Registrar, Academic Advising and the Comptroller’s Office. We will integrate document images stored on the system within the University’s administrative database system provided by CARS Information Systems. This will link them directly to the rest of the database information maintained on an individual or entity. This will also allow viewing of an image by any users connected to the University’s Ethernet Network, provided we grant view privileges to that person.

We need to accomplish this with minimal financial and staff resources by working cooperatively with existing vendors to benefit all parties.
I. BACKGROUND

St. Mary’s University is a comprehensive, private university in San Antonio, TX with an enrollment of approximately 4100 students.

Our Office of Financial Assistance processes Financial Aid application on approximately 3300 of those students each financial aid year. During a year the office will receive approximately 85,000 - 90,000 documents per year. These documents include things such as student and parent tax forms, financial aid transcripts from other institutions and financial aid verification forms.

Each year the financial aid office is spending $3500 -- $4000 dollars to have these documents microfilmed. In addition they keep approximately 350,000 to 400,000 hard copy documents in file folders stored within this office space.

II. PROBLEM

With limited physical space and the high demand for quick access to these documents the current manual filing system has become woefully inadequate to support the needs of the Financial Aid Office.

We needed to replace the current system with a technology solution in a short time frame and that fit into a limited budget.

III. APPROACH TO SOLUTION

We knew our solution would be to implement some form of on-line storage of the documents. The biggest drawback to this solution was being able to bring document imaging and storage on campus with the limited resources available.

I was sure that we did not want to bring in a totally new vendor into our current mix to handle this need. I have heard and read many horror stories about other institutions that have tried to effectively integrate document imaging systems into their existing networks and databases. In addition to the networking and software integration issues, We needed our document imaging solution to meet our work flow demands from a user interface point of view.

Our desire is to have an application that works the way our offices process the paper and to eliminate as may keystrokes and steps in the process as possible. The only solution to this was to leverage our long term relationship with our current administrative software supplier and work out an arrangement that was beneficial to both parties.

During the CAUSE94 conference I discuss this issue with our administrative software provider CARS Information Systems. We began exploring the possibility of a joint project to introduce document imaging into their product.

I was willing to commit a reasonable amount of our University’s resources to the project. This project would produce the basis for a generic document imaging application that integrates with our existing software systems and will be used by all administrative
operations with the need for document management. In addition to a limited monetary commitment the University would provide the necessary user input into the design and specifications for the application. We also would commitment to purchase the necessary equipment and to allow our vendor access to that equipment for development purposes.

The biggest hurdle to overcome at most institutions for this type of project is to secure the funds and authorization for this project from the University’s Executive Council and the President.

The first step in this process was to work with our vendor and identify as close as possible what the costs would be for this project. Internally I began to meet with the Director of Financial Assistance and the VP of Enrollment Management to determine just what they could afford to do within current budgetary constraints. We also took this opportunity to reiterate that this project would have a long term benefit to all administrative operation within the University. This was a very important step in laying the ground work for project approval with the Executive Council.

Our discussions with the Director and VP helped identify the fact that they did have money set aside for the purchase of a new filing system and for micro filming of existing documents. We made the determination to redirect a portion of those funds to this project. The Administrative Computer System budget would be able to provide $10,000 from moneys ear-marked for special development projects that were still available.

Once we determined what the available resources were and established strong support for the project across the University, we worked in cooperation with our vendor to establish costs for the project. That cost including hardware, software and miscellaneous expenses came to $33,600 (see appendix A). The funding breakdown is a follows; $10,000 from the Administrative Computer Systems budget, $2,000 from the current Financial Aid micro filming budget and the balance of $21,600 from a pool of excess revenues over expenditures at the end of the fiscal year. This money was approved by the Executive Council after we presented the overall benefits of the project.

Our vendor was very open to this cooperative approach and understood our budgetary limitations. They also were receiving many inquiries from other clients about imaging application and therefore had reason to want work cooperatively with a client on the initial application.

IV. PROJECT OVERVIEW

The initial objective of the Document Imaging Project will be to provide the Office of Financial Assistance with the ability to scan, store, retrieve, view and print the images of documents received by that office. This capability will provide 3 main benefits to the Office of Financial Assistance.

1. Saving of Space, the scanning and storage of document images electronically would eliminate the need to file and store the hard copy document within the small confines of the Office of Financial Assistance.
2. The ability to retrieve the documents of a particular student would be easy, fast, and available to multiple people simultaneously. Also we eliminate the time spent searching for misplaced or misfiled folders.

3. The ability to make backup documents and store the backup images at an off site facility. Most actual documents are not backed up because of the cost and amount of space required to store them.

The overall objective of this project is to provide the ability of storing documents received by any department within the University. Those departments would include but not be limited to, Admission, University Relations, Registrar and the Comptrollers Office. We will integrate the document images stored on the system within the University’s administrative database system, thereby linking them directly to the rest of the database information maintained on an individual or an entity. This would also allow viewing of any image by any users connected to the University’s Ethernet Network provided we grant access to that person.

V. INITIAL DELIVERABLES:

The deliverables agreed to between St. Mary’s and CARS Information Systems was to provide a system to handle the University’s Office of Financial Assistance documents. This system will function in the following manner.

1. As we receive documents, we collect them and take them to the scanning workstation. Here the operator will scan the documents and add indexing information via the workstation software.

2. We will return the documents to the person who will actually deal with them first. This person will retrieve the documents and verify image quality prior to the data from the document being entered into the system.

3. If the document later needed to be viewed or printed, then any user who has a networked PC connected to the CARS system will be able to call the document up and view or print it.

VI. PREPARATION AND SETUP ISSUES:

The introduction of document imaging systems into the work environment requires those persons using the system to change the way the think about working with paper documents.

One of the most difficult concepts to get through to staff who will be utilizing the application is to resist the desire to print a hard copy of the document. A recent article in Forbes ASAP magazine raises the issues about how document imaging systems can actually increase the amount of paper used within an operation. In that article listen to Walter Jacobs, an engineer installing Lotus Notes groupware for James River’s Pennington, Ala., production plant:
“Human beings trust the printed word,” Jacobs says, “that's something you'll never get away from.”

We found it necessary to work on changing these ingrained habits in order to achieve the maximum potential from an imaging system.

In introducing document imaging into an office you must work at changing the way in which staff work with information contained on these documents. Discussion about how documents will flow through the office are essential prior to the actual implementation of an imaging application.

We found that determining what is an acceptable quality of the document image is also an extremely important part for the setup and preparation of document imaging systems. This question relates directly to the issue having a cost effective imaging application. Although the cost of storage media is rapidly decreasing, the determination of what is an acceptable image to store weighs heavily on the amount of storage space you will require. A document scanned at 600 dpi will require six times the space as a document stored at 100 dpi.

Not only will the image cost more to store but it will also cost more in terms of the time needed to actually scan, store, retrieve and print the document. You must weigh this carefully when trying to keep the cost of imaging to an acceptable level.

When preparing to implement an imaging application set aside time to identify the many documents that you will be scanning. By setting up tables that identify the size, number of pages and the image quality setting for each individual document, you will save time during the actual act of scanning the documents to be stored.

VII. TOOLS AND TECHNOLOGY:

We are not purchasing a packaged imaging system, but instead we are having the application developed by our administrative software vendor. I thought it would be beneficial to describe the software and hardware that we will use in this application.

On the software side CARS Information Systems is using Visual Basic as the proto-typing tool. Eventually this will probably be replaced by Visual C++. Integrated Business Applications Builder is being used to develop the GUI front end and for data access. The actual imaging and scanning libraries are from Data Techniques, utilizing Intersolv's Data Direct ODBC Drivers, specifically for Informix 5. We should note that all development is being done with OLE compatibility in mind.

On the hardware and networking side our Financial Aid office is connected to the university’s Ethernet network and operates on a local Novell server. All workstations are
486 DX2 66mhz PC’s with 8mb of memory. They utilize Novell’s Lan Workplace to connect to our UNIX based administrative system.

The scanning workstation is a Pentium 100mhz PC with a 1.2 gigabyte hard drive and 16 megabytes of RAM. This machine is running Window for Workgroups and contains a SCSI host adaptor. We have connected a HP Scanjet 3C scanner and HP Laserjet 4M Plus printer to the system. We also have purchased a HP 20XT Optical Jukebox drive that contains 16, 1.2 gigabyte optical disks that is used to store the images. This drive will initially be connect to our host HP 847 administrative system but may eventually be moved to the local Novell server.

One of the early issues that we needed to address when planning this project was where, and in what manner, to store the scanned images. Our initial thoughts were to store the images in a database BLOB field. Our investigation of this option uncovered some very valuable pieces of information. First, in order to store the images within the database we would have to purchase additional software from Informix our database software vendor.

This software, *Informix Optical* presented a number of problems for us. First it was moderately expensive at a price of about $4000.00. Second this software would only support Write Once Read Many (WORM) access to the optical disk drive even though the disk itself was rewritable. Lastly we learned that Informix has no current plans to support the writing to CD-ROM in the near future. This was very important to us for the future since this likely will be the least expensive storage option that we will have in the future.

As we looked at our options we decided that the better direction to take was to store the images on the optical jukebox under the UNIX file systems. We will include just the indexing and pointer information in our database. In addition to providing us more flexibility we also feel that this will serve to increase the access time to the images and reduce the impact on the performance of our database.

**VII. ACTUAL COSTS INCURRED:**

One of the major reasons that St. Mary’s was able to move ahead with this imaging project was the relatively low investment that we needed to put into it initially. This combined with the gains in productivity that we expect to achieve from the project make us believe that we made a smart decision in pushing for the funding of the project.

The final initial startup costs broke down in the following manner and you can compare them to the initial estimates as outlined in Appendix A.

- Imaging Workstation Hardware
  - Pentium 100MHZ PC $ 3,026
  - 16mb RAM, SCSI Adaptor, 1 Gig Hard Drive
  - HP Laserjet 4mplus $ 1,812
  - HP Scanjet 3C Scanner $ 943
  - Auto Document Feeder $ 471

- Pentuim Fileserver for Financial Aid Office $ 2,639
APPENDIX A

INITIAL HARDWARE AND SOFTWARE REQUIREMENTS AND BUDGET

The following outlines the Hardware and Software requirements for the initial phase of the Document Imaging Project. Outlined here are the minimum requirements.

- **Imaging Workstation to include (REQUIRED)**  
  1. Pentium PC, 90mhz, 16mb RAM, 1 GB hard drive,  
     CD-ROM Drive, Network Card, 2mb Video, SCSI Card.  
     $3,500
  2. Mid range simplex scanner, monochrome,  
     accommodates 8-12 documents per minute  
     $2,000
  3. HP Laser Jet 4M Printer.  
     $1,500

- **Imaging Engine -- (REQUIRED) For initial trials the recommendation is to utilize a software engine. However, depending upon document flow and volume a hardware engine may be warranted, especially if a faster scanner is utilized. Note these hardware engines do not add any necessary functionality, just increase scanning speed.**  
  N/A

- **CARS Provided Software -- (REQUIRED) including the**  
  $12,000

We have brought this project in under the anticipated budget but actually spent the money on different items than was initially anticipated. You must also keep in mind that our Financial aid office had 486 DX2 PC’s on the essential staffs’ desktops prior to this projects undertaking. Obviously this would have increased the cost for this application had we had to purchase the workstations needed for viewing the stored images.
workstation imaging software and the add-ons to the CARS platform necessary to index the images in the database.

- Miscellaneous Software Licenses -- (REQUIRED) for scanning workstation, software for printing from any station and operating system software. $2,000

- Optical Disk Drive 20 GB HP Erasable Optical Jukebox Model 20XT (C1100A). One System provides ability to store 200,000 Documents on-line $7,100

- Informix On-line Optical Software to allow access to the optical disk drive. This software require a $825 annual maintenance cost. On the HP E55 maintenance is $510 annually $4,000

- Travel Expense for Training by CARS $1,500

TOTAL PROJECT COST $33,600

FOOTNOTES AND REFERENCES

Footnotes:


References: