DOCUMENT RESUME

ED 368 335 IR 016 575

AUTHOR Schwartz, Stanley F.
TITLE Electronic Document Imaging and Optical Storage Systems for Local Governments: An Introduction. Local Government Records Technical Information Series. Number 21.
INSTITUTION New York State Education Dept., Albany. State Archives and Records Administration.
PUB DATE 93
NOTE 24p.; For related document, see IR 016 576.
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; Archives; *Computer Storage Devices; Documentation; Glossaries; *Information Storage; *Local Government; Optical Disks; Planning; *Program Implementation; *Records Management; Technological Advancement
IDENTIFIERS *Electronic Document Imaging; New York

ABSTRACT

This publication introduces electronic document imaging systems and provides guidance for local governments in New York in deciding whether such systems should be adopted for their own records and information management purposes. It advises local governments on how to develop plans for using such technology by discussing its advantages and disadvantages and its relation to records management. Electronic document imaging systems are computer-based systems that capture images of paper documents through a scanner, index them to expedite retrieval, and store them on magnetic, magneto-optical, or optical media, in or near the computer. Electronic document imaging systems are not panaceas, but they can offer many advantages to a local government in terms of space saved, rapid and easy access, and the improvement of delivery, distribution, and security. This overview highlights records management issues such as hardware and software, records retention, disaster planning, authentication, system documentation, security, and legal issues. A glossary of imaging terms explains many common aspects of electronic document imaging. (SLD)

IR

21

Electronic Document Imaging and Optical Storage Systems for Local Governments: An Introduction

Local Government Records
Technical Information Series

NEW YORK
THE STATE OF LEARNING

The University of the State of New York
The State Education Department
State Archives and Records Administration
Local Government Records Services
Albany, New York 12230

NUMBER

21

Electronic Document Imaging and Optical Storage Systems for Local Governments: An Introduction

Local Government Records
Technical Information Series

Stanley F. Schwartz
1993

The University of the State of New York
The State Education Department
State Archives and Records Administration
Local Government Records Services
Albany, New York 12230

THE UNIVERSITY OF THE STATE OF NEW YORK
Regents of The University

R. CARLOS CARBALLADA, *Chancellor*, B.S.	Rochester
JORGE L. BATISTA, *Vice Chancellor*; B.A., J.D.	Bronx
WILLARD A. GENRICH, *Chancellor Emeritus*, LL.B.	Buffalo
EMLYN I. GRIFFITH, A.B., J.D.	Rome
LOUISE P. MATTEONI, B.A., M.A., Ph.D.	Bayside
J. EDWARD MEYER, B.A., LL.B.	Chappaqua
FLOYD S. LINTON, A.B., M.A., M.P.A.	Miller Place
MIMI LEVIN LIEBER, B.A., M.A.	New York
SHIRLEY C. BROWN, B.A., M.A., Ph.D.	Albany
NORMA GLUCK, B.A., M.S.W.	New York
ADELAIDE L. SANFORD, B.A., M.A., P.D.	Hollis
WALTER COOPER, B.A., Ph.D.	Rochester
CARL T. HAYDEN, A.B., J.D.	Elmira
DIANE O'NEILL MCGIVERN, B.S.N., M.A., Ph.D.	Staten Island
SAUL B. COHEN, B.A., M.A., Ph.D.	New Rochelle
JAMES C. DAWSON, A.A., B.A., M.S., Ph.D.	Peru

President of The University and Commissioner of Education
Thomas Sobol

Executive Deputy Commissioner of Education
Thomas E. Sheldon

Deputy Commissioner for Cultural Education
Carole F. Huxley

Assistant Commissioner for Archives and Records Administration
Larry J. Hackman

Director, External Programs
Bruce W. Dearstyne

Chief, Local Government Records Services
Robert W. Arnold III

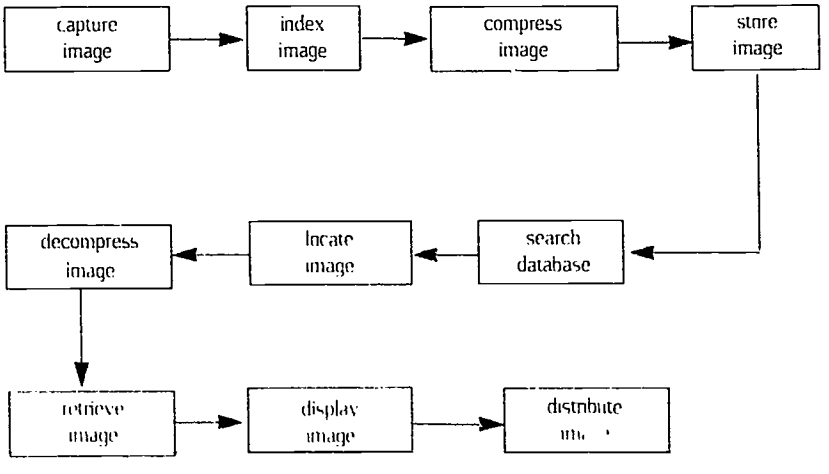
Electronic Document Imaging and Optical Storage Systems for Local Governments: An Introduction

Local Government Records Technical Information Series No. 21

This publication introduces *electronic document imaging systems* and provides guidance for local governments in deciding whether such systems should be adopted for their own records and information management purposes. It is issued by the State Archives and Records Administration as part of its statutory responsibility to provide advice and guidance to local governments on records management, including the adoption and use of information technology to enhance records management. For a more in-depth discussion of optical storage, consult Series No. 45. It advises local governments how to develop plans for using such technologies by discussing their advantages and disadvantages and how they relate to records management.

Electronic document imaging systems are computer-based systems which capture images of paper documents through the use of a scanner; index them to expedite retrieval; and store them on magnetic, magneto-optical or optical media, housed either inside or near the computer. Future publications will examine aspects of electronic document imaging technology, management issues associated with it,

Electronic Document Imaging System



records management considerations raised by it, and local government uses. This overview discusses each of these issues in the broadest, least technical terms.

Organizations use electronic document imaging systems to improve their performance on a variety of tasks. In some cases, these systems are considered an alternative to paper or microfilm storage and retrieval. Electronic document imaging systems are not panaceas for all records management problems but, in an appropriate application, they offer significant advantages in meeting a government's records management objectives. Such systems may yield savings in records storage space; reduce the time needed to access and retrieve documents; and improve document delivery, distribution, and security.

Optical Storage Technology

In the most general terms, images of documents are "recorded" through a *scanner.* Scanners look and function like photocopiers, but instead of copying onto paper, they "copy" images of documents into a digital format that is stored on a computer-readable disk or tape. Scanners are discussed below in the section on "hardware." While the images scanned into a system may be stored on either magnetic or optical computer storage media, they are most commonly stored on optical media because the files are large, and optical media have extremely high storage capacity. During the process of recording images, the images are indexed to expedite future retrieval. These indexes are maintained in a database, which may be stored with the images on optical media. However, indexes are most often stored separately from the images, often on magnetic media.

When optical disks are "written to" or "recorded," lasers alter their surfaces, thereby changing their light reflecting properties. These alterations are "read" by software during the retrieval process. Once retrieved, images may be displayed on a monitor, or printed or plotted on paper, or copied onto microfilm. These images may also be distributed via telecommunications networks, so that several users may see and act on them simultaneously.

Optical disk is the most common optical storage medium. An optical disk is platter-shaped and housed in a plexiglass or plastic container. Optical disks used by governments for storing computer data are generally manufactured in two types: *rewritable (erasable)* and *write-once (nonerasable),* sometimes called *WORMs (write once, read many [times]),* and should be distinguished from the more common Compact Disc-Read Only Memory, or CD-ROM which is used primarily as a distribution medium for a variety of publications. While the nonerasable type of optical disk seems ideal for storing public records, this type is losing favor with manufacturers. Market forces have worked toward increasing the number of rewritable disks man-

ufactured because they are far more economical to use. Because erasable disks may be altered or erased, using them to store the only copy of public records is not recommended. However, it should be noted that all magnetic storage media are erasable and that records stored on any computer media may be protected to a great degree by developing, implementing, and documenting procedures for recording, storing, maintaining, accessing, and changing records. Following such procedures will reduce the negative impact of using rewritable disks.

Sections below focus on legal issues related to optical storage, including choosing between WORM and rewritable formats. Other considerations in choosing optical disk formats include the longevity and stability of the media, and predicting the likelihood of the chosen format's continuing availability.

Indexing

Indexing may be the most critical component in electronic document imaging systems. The purpose of the index is to quickly lead the user directly to the image being sought or, if that is not possible, to a set of images that include the image being sought. Indexing decisions, such as the number and type of fields included in the index, will have an impact on both how quickly records are entered into the system and how accurately records are retrieved from the system.

Unless another approach is justified, indexing should be done on a small number of fields and require as little interpretation as possible. Such fields as names, case numbers, social security numbers, section, block and lot numbers, and other specific information items, are better choices than subject headings, which may be open to interpretation or lead to confusion. Indexing on a small number of easily defined fields is easier to do for some types of documents than for others. In the case of records such as minute books where subject searching is likely, index fields should be limited, and terms should be carefully controlled through the use of a thesaurus or other authority control technique.

Optical Disk Systems: Advantages and Disadvantages

Many people have come to equate storing images of paper documents on optical disk with the process of electronic document imaging. Optical disk systems are often included in electronic document imaging systems because they offer a great deal of storage capacity in a relatively small space.

At their most elemental level, images stored on optical disk are a substitute or surrogate for paper or for microfilm. Many smaller, less expensive imaging systems are designed specifically for this purpose. These systems allow for image capture, indexing, and storage of images. Like microfilm, they have the advantage of reducing the physical space required for records storage. Like those microfilm

systems enhanced with computer assisted retrieval (CAR) software, these systems can quickly lead searchers to the images they seek. Electronic document imaging systems are superior to microfilm during the image capture stage of the process because these systems allow the user to see immediately whether the captured image is usable and do not require professional processing prior to their use. Another advantage of electronic document imaging systems is the ready distribution of images made possible because digital files are created when images are scanned. These files may be saved and distributed like any other digital file, infinitely increasing the possibility of sharing images over telecommunication systems.

There are also disadvantages to using optical media, especially for the storage of records scheduled for retention for long periods of time. The longevity and stability attributes of optical disks are still not well-known; it is not always possible to determine, accurately, the recording life and the retrieval life of specific optical disks. Estimates provided by manufacturers are based on numerous assumptions of doubtful reliability. While the longevity and stability of optical media could be problematic, they cause less concern than the longevity and stability of electronic document imaging and optical storage technologies themselves.

To better understand one important aspect of the technology problem, think about such technologies as videocassette recorders. Less than a decade ago, there seemed to be a clear choice between Beta format and VHS. Consider the problems faced by a government whose officials had chosen Beta and which now needs to replay a tape recorded in that format. Another example may be useful. Suppose a government had stored records on computer punch cards or on an older format magnetic tape. It would be difficult, maybe even impossible, to access those records today. The problem is not that the cards or tapes would have deteriorated. Rather, the equipment required to read their contents has disappeared from most work environments.

At present, there are few standards to assure governments that images of documents captured today will be accessible in five years. This is a common problem with optical disk and all other machine-readable (rather than eye-readable) media. The solution rests in frequently and regularly copying images to fresh media capable of being read by new equipment. However, this solution assumes that future funds will be committed to this activity. This is not always a valid assumption. Given these considerations, records with a retention period of 10 or more years that are stored on optical disk should be kept on paper or on microfilm, and the paper or microfilm should be considered the record.

A government considering using optical disk to store public records should carefully study the scope of its project and anticipate its future requirements. Important decisions include whether the conversion

from paper to digital images will be done retrospectively, that is, imaging the entire or some significant portion of the "back file" of the records series, from the day the new system is installed, or in some other manner. A government could have a policy under which all new cases are imaged and the back file of any reopened existing case is also imaged. If a retrospective conversion to the new system is planned, the government must decide whether it will be done in-house or outside, and whether this activity will be done by government employees or by contractors. These decisions carry with them significant cost considerations. Other important decisions include who will have access to the image database and the images themselves, how searches will be conducted, and how images will be distributed and used.

Electronic Document Imaging and Work Flow

Using optical disk to store public documents may benefit governments primarily by decreasing records storage space requirements and improving access to records. However, many electronic document imaging systems, which promise the greatest return on investment, emphasize changes in the way work is done and the way information is processed. Applications which emphasize "workflow" redirect or speed up the ways paper and information move through a system and how they are handled at each point in a process, as well as how they are stored and retrieved. Such workflow systems are designed to improve performance on work tasks and provide new structure to how work is done. Some corporations, educational institutions, and governments have identified tasks that are improved by workflow based electronic document imaging systems. Some county clerks' offices now capture, store, and retrieve deeds and mortgages. Other uses might include processing permits or licenses, or streamlining business filings or case files.

In general, activities that are especially well-suited to the adoption of workflow based electronic document imaging systems include the following:

- transaction-oriented tasks, often containing numerous documents emanating from different sources at different times and needing to be linked with each other
- events and activities needing to be tracked
- high, relatively steady, not seasonal, volume of activity
- records series with short, but intense, active reference periods including documents that rarely require physical changes or editing

Workflow based systems improve efficiency, saving time, money, or both. In these systems, both old and new images may have to be

searched and retrieved. For example, during a title search, older images will be called up. Case files, especially in social services departments, move quickly from inactive to active status. Such searching and retrieving frequently calls for quickly accessing documents, which may be stored on a number of different optical disks. Some systems often store disks together in an autochanger or "jukebox." These devices may hold a number of disks of the same or of different types and, in the latter case, disk drives on which information may be retrieved.

In addition to providing the capability of entering public records quickly and easily into a computer system, electronic document imaging rapidly indexes them to be conducted in a variety of ways and from locations which may be some distance from the government offices. In some systems, images of documents may be faxed to a searcher or downloaded directly to the searcher's computer.

System Planning and Implementation

Needs Assessment and Feasibility Study

Any government or agency should proceed with caution before making decisions about implementing an electronic document imaging system. The key to a successful implementation is careful study of the intended application and the relationship of this application to the other information resources held by the organization. The first step is to state clearly the problem to be solved. Determine and enumerate the benefits to accrue from the implementation of this technology. In other words, it is important to state a vision of how the implementation of the system will improve performance and have a positive impact on service delivery. These steps will assist a government in justifying the costs of implementing this technology. A needs assessment and feasibility study should analyze the work that is being done, how, and how well tasks are accomplished, and the capabilities and work experience of people currently doing the work. Before deciding to solve the problem by implementing an electronic document imaging system, the needs assessment should identify a variety of manual and technological approaches which may be used to efficiently and effectively complete tasks.

The needs assessment should result in the identification of a clearly defined problem, a thoroughly reviewed and assessed workflow (important even for those applications which will essentially store images, since the work required to capture, index, store, and search those images needs careful planning and staff training), and a statement of the costs associated with the project, including hardware and software acquisition, staff training, running dual systems while the new system is being debugged, etc. The needs assessment/feasibility study should also identify other governments addressing this problem and determine how they have approached or solved it. If other governments have implemented a technological solution, the feasibility

study should determine and report on what has been done, how long the solution has been in place, whether it has met goals and objectives, what the significant problems have been, the initial and continuing costs, including training and maintenance, and pitfalls to avoid.

In many cases, technology-based solutions found effective for governments addressing specific problems will also work for similar governments addressing these problems. Good managers will determine similarities and differences and draw conclusions about the appropriateness of solutions implemented elsewhere for solving their own problems. The guidelines set forth in the Association for Information and Image Management *Resource Report: Feasibility Studies for Document Management Systems* are very useful (available from AIIM, 1100 Wayne Avenue, Suite 1100, Silver Spring, MD 20910).

After assessing needs and determining feasibility, if a government decides to pursue an electronic document imaging system, it will usually write a Request for Proposal (RFP), a formal document written to present to potential contractors a set of requirements and specifications to be met by a proposed system. Another AIIM publication, Technical Report TR27-1991 *Electronic Imaging Request for Proposal (RFP) Guidelines*, provides assistance in writing an RFP. Often, a consultant is hired to conduct the needs assessment and feasibility study and to write the Request for Proposal.

Needs assessments, feasibility studies, and preparing requests for proposals may be done by government staff or by consultants. Many system vendors offer consulting services in this area, and it is sometimes difficult to separate services from sales. Government officials should be aware of the status of the individuals or firms with whom they are working, and the expectations of each party should be made clear in letters of understanding or in contracts.

Considerations in Selecting Electronic Document Imaging Systems

Electronic document imaging systems contain many hardware and software components. The computer and all of the peripheral devices must be chosen carefully to meet the specifications of the system design. However, the most critical decision is the selection of software to handle image capture, compression, enhancement, storage, linking, indexing, searching, retrieving and displaying, and incorporating images with existing databases. Software choices will dictate the success or failure of a project.

After determining its requirements, a government should arrange for demonstrations by different systems manufacturers or vendors to see how their software handles each task. Commercial software should be assessed in the context of the government's needs and specifications.

If the project can proceed using "off-the-shelf" software, this will control costs throughout the life of the project and allow the government to take advantage of upgrades and improvements developed by the manufacturer. While such an approach may be ideal, the project may require a higher degree of customization. One approach is to customize the off-the-shelf software. In this approach, the manufacturer or a company working with the manufacturer customizes the software to meet a government's requirements. This assumes that a long-term relationship will be established for continued software maintenance and support.

If customizing off-the-shelf software does not meet the government's requirements, a third approach is to have a contractor create software to a government's exact specifications. This is not only very expensive, but also risky because the government will continue to depend on the manufacturer for the life of the product. Because the imaging business is still new, and as in any new field there are business failures or corporate decisions to leave the business, SARA highly recommends that governments require vendors to provide the source code (the actual programming) along with the software. At the very least, a government should arrange for the source code to be deposited in a bank vault and make legal arrangements to have access to the source code in the event that the manufacturer ceases its involvement with imaging or ceases to exist.

Records Management Issues

Nonproprietary Hardware and Software

The key records management issues for electronic document imaging systems are effective and efficient access and retrieval of records throughout their retention period. These issues are related to the selection of storage media, file formats, and system components. It is most important to avoid systems using proprietary components or software. Proprietary components are those which are the exclusive property of a manufacturer and are used by that manufacturer to differentiate its products from competitors' products. Vendors may argue that their proprietary systems improve performance. This may be true, but the ongoing effective use of a system, especially in the absence of its creator, is far more critical to its overall success than most proprietary improvements, real or imagined.

Retention Periods

Electronic document imaging systems are not considered reliable for local government records with a ten-year or longer retention period. The reason is that there are no reliable standards for the longevity of images on such systems, and the hardware and software may become

obsolete after several years. For records with longer retention periods, local governments should retain a paper or microfilm copy to ensure that the information will survive and be available for as long as legally required.

Disaster Planning

Electronic document imaging systems require careful disaster planning. As more and more government records are being created and stored in computer systems, the meaning of the phrase "disaster planning" has been broadened. Of course, fire, flood, and other natural disasters can affect electronic records as they do those kept on paper or microforms. However, electronic records are also vulnerable to power surges, power outages, computer viruses, and other disasters. Disaster planning now encompasses steps taken to ensure the maintenance of data and data integrity. In the case of electronic document imaging systems, such plans include routine and frequent backup of all image and index files and storage of both at remote sites.

Authentication

An authentication plan provides local governments with the basis for establishing the credibility and legal admissibility of records stored in electronic or photographic formats. The government should create an authentication plan that includes a written policy that describes how the electronic document imaging system is used in the "normal course of business," identifies the records series processed in the system, and links the system and the images to approved retention schedules. The plan should also include the name of a system administrator, an individual who is knowledgeable about the system and its operation and who can be called upon to explain or to testify about what the system does and how it works.

System Documentation

Documentation of system characteristics includes code for system software and application software. Hardware, including the configuration of the computer and the peripherals, should also be described in detail. Operating procedures should be listed and described. This description should address the following steps:

- preparation
- scanning
- indexing
- creating and updating the index database
- creating and updating the image database
- enhancing and manipulating images (if done)

- compressing and decompressing images
- networks used
- search and retrieval procedures and software
- displaying images at the workstation, including specifications for display devices
- printing

Each step should be described and the names of individuals responsible (where applicable) should be listed for each step.

Security Considerations

Security is critical for any computer system in which public records are created or kept. All activities associated with electronic document imaging systems must be documented. Lists of users, access practices and privileges, including the use of protected passwords and other security measures, must be carefully kept. System and manual logging of activity, such as changes to records, are also required. These considerations should be discussed at length with vendors.

Legal Issues

At the time of this writing, the State Archives and Records Administration is not aware of any case law establishing the admissibility of images of documents stored on any electronic media. However, the logic established in the *Uniform Photographic Copies of Business and Public Evidence Act (UPA)*, though passed in 1949 and therefore not addressing specifically images stored on optical disks, seems to point to these images as "true copies" of documents. A "true copy" is considered an adequate substitute for the original, for all purposes that the original was created or maintained. The images stored on optical disks are subject to laws pertaining to copies. They are also subject to the "best evidence rule." This rule allows for the submission of the best available evidence. If paper originals or copies have been destroyed in the normal course of business and in accordance with the law, and images stored on optical disk are considered the best evidence available, in all likelihood they will be accepted.

Public records of all types must satisfy legally mandated retention periods and legally mandated access requirements and be admissible in evidence. Images stored on optical disk appear to meet the legal requirements discussed above. However, given the rapid changes found in this technology, images of public records scheduled to be retained for more than 10 years should not be retained only in electronic form. As noted above, paper or microfilm should be used to meet these longer retention requirements. Each government should

consult its legal counsel to discuss these issues before purchasing an electronic document imaging system.

A final legal requirement for images stored on optical disk may be their court-ordered expungement. Frequently, social service agency records and police records documenting the activities of juvenile offenders fall into this category. WORMs cannot be expunged because images are permanently recorded on them. However, WORMs can be *virtually* expunged by destroying the indexes to the images. In some cases, virtual expungement may not satisfy court orders. It is possible to reproduce a WORM without producing those sectors containing the expunged data. This procedure could become quite expensive if expungement is regularly required. Of course, the contents of rewritable disks can be expunged. Using rewritable disks, however, leads to a set of problems which are not found when using WORMs, since they may be more easily altered. One solution is to use rewritables for expungeable records and WORMs for all others. This solution requires the purchase of two different optical disk drives and will have an impact on how work is planned and conducted.

Costs

As may be expected, costs vary significantly. Costs depend on the size and capacity of the system, the nature of the computer, peripheral devices, storage, indexing techniques, etc. At this writing, it is possible to purchase a complete stand-alone system, including a computer, software, scanner, monitor, and a single optical disk drive for less than $20,000. Other systems may be put together (either by internal staff or by a systems integrator) using computers already owned by a government or by adding servers and peripherals to local area networks (LANs). Fully functional workflow systems serving an entire enterprise remain very expensive (often into the millions of dollars) and generally require custom programming, considerable design, and installation time, effort, and staff training. Justifying any of these systems presents a challenge to most governments.

For More Information and Assistance

Another technical publication, #45, *Optical Storage Systems for Records and Information Management: Overview, Recommendations and Guidelines for Local Governments*, is also available from SARA.

The State Archives and Records Administration provides records management services to local governments including technical advice and assistance, publications, training and presentations, and consultations with local government officials concerning records and information management issues. SARA has regional offices throughout the State; each office has an expert records specialist who

can visit local governments and provide on-the-spot advice. These services are supported by the Local Government Records Management Improvement Fund. For further information, contact your regional office, or—

Local Government Records Services
State Archives and Records Administration
State Education Department
10A63 Cultural Education Center
Albany, NY 12230
(518) 474-6926

Glossary of Imaging Terms[1]

Algorithm	Prescribed set of mathematical steps which is used to solve a problem or conduct an operation.
ANSI	American National Standards Institute. A standard-setting, non-government organization, which develops and publishes standards for "voluntary" use in the United States.
API	Application Program Interface. Generic term for any language and format used by one program to help it communicate with another program. Specifically, an imaging vendor can provide an API that enables programmers to re-package or recombine parts of the vendor's imaging system or integrate the imaging system with other applications, or to customize the user interface to the imaging system.
ASCII	American Standard Code for Information Interchange. Pronounced "Ask-ee." The most popular coding method used by small computers for converting letters, numbers, punctuation and control codes into digital form. Once defined, ASCII characters can be understood by other computers and by communications devices. ASCII represents characters, numbers or punctuation marks in seven on-off bits. A capital "C" is represented as 1000011, while a "3" is 0110011. This allows all PCs to talk to each other if they use a compatible modem and transmit at the same speed.
Authorization code	Identifying code, often a password, that allows a user access to a system. Used mainly for privacy and security.
Autochanger	Device that holds multiple optical disks and one or more disk drives, and can swap disks in and out of drives as needed. Same as "jukebox."
Auxiliary storage	External storage devices, such as disk drives, optical drives and tape drives and the media used on them.
Backbone	Part of a communications network which carries the heaviest traffic. Part of a network which joins LANs together. LANs are connected to the backbone via bridges or routers and the backbone serves as a communications highway for LAN-to-LAN traffic.
Backfile conversion	The process of scanning in, indexing and storing a large backlog of documents on an imaging system. Time consuming. Generally performed by a service bureau.

[1]An abridgement and adaptation of *The Imaging Glossary*: Electronic Document & Image Processing Terms, Acronyms and Concepts by Andy Moore. New York: Telecom Library. Telephone: 1 800 999 0345. Used with permission.

Bar code	A system of portraying data in a series of machine readable lines of varying widths. Used to encode indexing information.
Bit	Contraction for binary digit. The smallest unit of data a computer can process. Represents one of two conditions, on or off; 1 or 0, mark or space, something or nothing. Bits are arranged into groups of 8, called bytes. A byte is the equivalent of one character.
Bit map	Representation of characters or graphics by individual pixels, or points of light, dark or color, arranged in row and column order. Each pixel is represented by either one bit (simple black & white) or up to 32 bits (high definition color).
Bit specifications	Number of colors or levels of gray that can be displayed at one time. Controlled by the amount of memory in the computer's graphics controller card. An 8 bit controller can display 256 colors or levels of gray. A 16 bit can show 64,000 colors, a 24 bit controller can display 16.8 million colors or gray levels.
Cache	Pronounced "cash." Small portion of high speed memory used for temporary storage of frequently used data. Reduces the time it takes to access that data.
CCD	Charge Coupled Devices. A type of digital camera technology in which the image is focused on an array of sensing pixels. The small size of the array and the high resolution have enhanced image acquisition. Most scanners employ CCDs.
CCITT	Consultative Committee on International Telegraphy and Telephony. The source of the most commonly used compression algorithm in imaging systems.
Compression	A software or hardware process that "shrinks" images so they occupy less storage space and can be transmitted faster and easier. Generally accomplished by removing the bits that define blank spaces and other redundant data, and replacing them with a smaller algorithm that represents the bits. No matter how data is compressed, it must be decompressed before it can be used.
Contextual search	To locate documents stored in a system by searching for text that appears in them, rather than searching for them by file name or other indexing technique. Contextual searching cannot be done on images.
DASD	Direct Access Storage Device. Any on-line disc data storage device. Generally used to refer to internal magnetic storage in mainframe and mini computers.
DAT	Digital Audio Tape. A technology that records noise-free digital data on magnetic tape. Can hold up to 2 Gigabytes of data.
Density	Degree of darkness of an image. Also, the percent of the screen used in an image.

Departmental imaging system	A multi-workstation imaging system, typically used by at least three and usually more than 25 people in a workgroup or department.
Descriptor	The key word, code or phrase that an automated document retrieval system uses to identify and locate the document.
Desktop imaging system	An imaging system with a single workstation, often a microcomputer, meant to be used by one person at a time.
Device driver	Small program that tells the computer how to communicate with a particular type of peripheral device (a printer or a mouse or some other type).
Digital	The use of binary code to record information. "Information" can be text in a code such as ASCII, or scanned images in a bit mapped form, or sound in a sampled digital form, or video. Digital recording increases the ease of manipulating data and improves the accuracy of transmission.
Digitize	To convert an image or signal into binary code. Visual images are digitized by scanning them and assigning a binary code to the resulting data.
Disk drive	A device containing motors, electronics and other gadgetry for storing (writing) and retrieving (reading) data on a disk.
DPI	Dots per inch. A measurement of input or output device resolution and quality. Measures the number of dots a scanner can scan, a printer can print or a monitor can display horizontally and vertically.
Eye readable	Images which can be read by the human eye. In the case of microforms, eye readability requires magnification.
Fax	A collection of technologies in which paper documents are first scanned then digitized, then the digital image is converted to analog form. The fax machine dials and arranges a data communications session (agreeing on speed of transmission and protocol) with a remote machine, and transmits the analog document. The receiving machine captures the analog data, reconverts it to digital form and prints a copy of the original. Current fax machines are called Group 3. Group 4 fax will become commonplace when the public telephone system is all digital.
Feature extraction	An optical character recognition (OCR) technique in which the software keeps data regarding all characters' features, i.e. "the letter A has two diagonal lines; the lines intersect at the top; it has a horizontal line that crosses from one of the lines to the other, etc." As the OCR scans, it compares features to its feature library.
Frame grabber	A device that changes a video picture into a digital computer graphics language.

Full text search	The ability to search text files for occurrences of certain words, digits, sentences or patterns of characters. Scanned documents cannot be full text search. To do that, the document would have to be re-typed or the image put through an OCR process to create a text file.
Gigabyte	A million bytes or a thousand megabytes of data. Storage capacity of this size is needed to cope with the large files created in imaging systems.
Gray scale	The spectrum or range of shades of black an image has. Scanners' and monitors' gray scales are determined by the number of gray shades or steps they can recognize and reproduce. A scanner which can see a gray scale of 16 will not produce as accurate an image as one that distinguishes a gray scale of 256.
Handshaking	Exchange of signals at the beginning of a data communications session. During this exchange, the two systems confirm each other's specifications, including parity, baud rate and speed, to ensure that a proper link is set for the data transmission.
Handwriting recognition	A system for taking handwritten material generated with a stylus on a computer pad or directly onto the computer screen, and converting it to machine readable text.
Hard disk	A storage device that uses a magnetic recording material. Generally, hard disks are fixed inside a PC, but there are removable versions. Storage capacities vary.
Image	The computerized representation of a picture or graphic.
Image processing	The manipulation of digitized images which have been acquired through scanning or some other method. Image processing allows for the comparison and analysis of characteristics which the human eye alone cannot see.
Image resolution	The fineness or coarseness of an image as it was digitized, measured as dots per inch.
Imaging system	Collection of units that work together to capture and recreate images. At its simplest, it is an acquisition device (scanner), an image processor and an imaging device (printer, microform or monitor).
Index	A descriptive set of data associated with a document for locating the document's storage location. Indexing stored documents is the greatest challenge in document retrieval. It is difficult to devise an indexing scheme which describes all possible parameters for later searches, comparisons and processing. For most record series, this is not a goal.
Input workstation	The location at which paper or microform documents are scanned and computer files entered. Also the place where the index is assigned to the document.

JPEG	Standard for still image compression devised by the Joint Photographic Experts Group and sanctioned by the International Standards Organization (ISO) and CCITT. JPEG compression is used for photographic images, not for text images.
Jukebox	A device that holds multiple optical discs and one or more disc drives and can swap discs in and out of the drive as needed. Same as an autochanger.
LAN	Local area network. High speed transmission over twisted pair, coaxial or fiber optic cables that connect PCs, mainframes, terminals and peripherals together at distances of one mile or less.
Laser disc	An optical disc with the same technology as a CD, except laser discs are 12" in diameter.
Lossless	Image and data compression applications and algorithms that reduce the number of bits a picture would normally take up without losing any data.
Lossy	Methods of image compression (such as JPEG) that reduce the size of an image by disregarding some information.
MAN	Metropolitan area network. Usually involves high speed transmission over fiber optic cables that connect LANs together at distances of up to 25 miles.
Machine readable	Data in a format (i.e. ASCII) or on a medium (disks, tapes, optical discs, or punched cards) that a computer can understand.
Magnetic recording	A technique of recording analog or digital signals or data on a medium of specially prepared grains of iron oxide. Storage media includes hard disks, floppy disks, magnetic tape, and others.
Magnetic tape	Storage medium that uses a thin plastic ribbon coated with iron oxide compound to record data with electrical pulses. A sequential (serial) storage medium in which the next bit of data is recorded after the last and you have to look through the whole tape to find what you are seeking.
Magneto-optic	A high density, erasable recording method combining magnetic and optical technologies.
MPEG	An image compression scheme for full motion video proposed by the Motion Picture Experts Group. It notes differences between consecutive frames and only stores those which are different.
Node	A point of connection into a network.
OCR	Optical character recognition. The ability of a scanner with the proper software to capture, recognize and translate printed alphanumeric characters into machine readable text. OCR uses either "pattern matching" or "feature extraction." With pattern matching, the software has a template of possible characters. A letter is compared to a library of pattern

templates. If it is recognized, the ASCII equivalent of the letter is sent to the output file. "Feature extraction" is described earlier in this glossary.

Omnifont — The ability of an OCR to recognize any typeface font without having to learn it in advance. It uses feature extraction.

Optical disc — A direct access storage device that is written and read by laser light. Some optical discs are WORMs (write once, read many) on which data is permanently engraved in the surface in some way. Erasable optical discs are described below in this glossary. Compact discs and laser discs are optical discs which are used for distribution purposes and are rarely created one at a time.

In general, the storage capacities of optical discs are far greater than of magnetic discs, but their access times are much slower.

Optical scanner — Input device that translates human readable or microform images to machine readable data.

Packet — A group of bits, packaged together, for transmission purposes. Three principal elements are included in the packet: (1) control information, i.e. destination, origin, length, etc.; (2) data to be transmitted; and (3) error detection and correction bits. More efficient than sending information in continuous streams.

Page recognition — OCR software that can tell the difference between text on a page and other items such as pictures and art work.

Pattern matching — See OCR in this glossary.

Phase change recording — An optical recording technique.

Pixel — Picture element. When an image is defined by many tiny dots, those dots are pixels. On the printed page, each pixel is one dot. On a color monitor, each pixel can be made up of many dots.

Plotter — A printer that prints vector graphics, i.e. images created by a series of many straight lines. Used in geographic information systems and CAD.

Proximity search — A feature of full text searching in which every occurrence of a word within a certain distance of another word is found.

QIC — Quarter inch cartridge. A common format for magnetic tape for data storage.

Resolution — A measure of image output capability, usually expressed in dots per inch (DPI). The higher the resolution, the greater amount of detail that is shown.

Re-writable — Optical media from which data can be erased and new data added. Magneto optical and phase change are the two main types of re-writable optical discs.

RISC — Reduced instruction set computing. A system with a special micro-processor which processes fewer instructions and is very fast. RISC

	computers are often used in calculation-intensive operations including GIS and CAD.
Scanner	A device that optically senses a human readable image and contains software to convert the image to machine readable code.
SCSI	Pronounced "scuzzy." An industry standard for connecting peripheral devices and their controllers to a microprocessor. The SCSI defines the hardware and software standards for communication between a host computer and a peripheral.
TCP/IP	Transmission control protocol/internet program. A set of protocols developed by the Department of Defense to link dissimilar computers across networks.
Terabyte	A trillion bytes or 1,024 gigabytes.
TIFF	Tagged Image File Format. A bit map file format for describing and storing color and gray scale images.
UNIX	A general purpose, multi-user, multi-tasking operating system. UNIX is powerful and complex and needs a computer with a large amount of RAM. It works on different computers and when used wisely can make easier the process of moving programs from one computer to another.
Video digitizer	See "frame grabber."
Videodisc	Read-only, direct access optical disc that holds up to two hours of analog video data.
Workflow system	An imaging system that concentrates on automating the processing of documents, i.e. scheduling processing, routing documents among departments and tracking document status.
WORM	Write once read many. Optical storage device on which data is permanently recorded. Data cannot be altered.
X.25	A standard recommendation from CCITT. X.25 has come to represent a common reference point by which mainframes, minis, micro computers and specialized terminals from many manufacturers can be made to work together over a packet switched network.

The University of the State of New York
The State Education Department
State Archives and Records Administration
Local Government Records Services
Albany, New York 12230