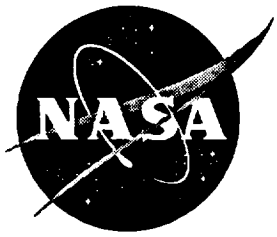


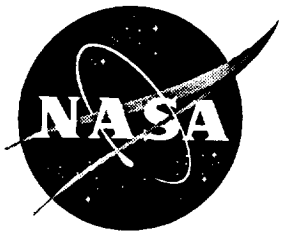
NASA Contractor Report 198756

CD-ROM Preparation: An Overview and Guide

Ralph E. Daniel, Mark W. Jeschke, and James A. Schroer

Contract NASw-4584
Prepared for NASA Scientific and Technical Information Office

June 1995



NASA Contractor Report 198756

CD-ROM Preparation: An Overview and Guide

Ralph E. Daniel, Mark W. Jeschke, and James A. Schroer
RMS Associates • Linthicum Heights, MD

National Aeronautics and Space Administration
Center for AeroSpace Information•Linthicum Heights, MD

Prepared for
NASA Scientific and Technical Information Office
under Contract NASw-4584

June 1995

This publication was prepared by the NASA Center for AeroSpace Information,
800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934, (301) 621-0390.

Acknowledgements

This work was funded under the NASA Scientific and Technical Information Office contract with the NASA Center for AeroSpace Information, NASw 4584. The authors wish to extend their thanks to Ms. Terese Ohnsorg of the NASA Scientific and Technical Information Office for her support and helpful suggestions during the preparation of this manuscript.

CD-ROM Preparation is the resulting document from TD 94-031. This project has two objectives:

Prepare a primer of the options and procedures involved in producing a CD-ROM.

Prepare an overview of the marketplace, with an emphasis on equipment and software availability.

This report will stand alone as a roadmap to producing a CD-ROM product.

1 CD-ROM Technology—Introduction

CD-ROM is a permanent optical storage device. Linked to a PC, it becomes a powerful peripheral, putting millions of bits of data at the user's fingertips. While CD-ROM technology is less than ten years old, it is a rapidly emerging field with new applications being identified and new commercially available hardware and software being produced every day. CD-ROM stands for Compact Disc Read-Only Memory. This means that data can be stored and accessed but not edited. CD-ROM discs are read optically by a laser beam, similar to the way an audio compact disc is played on a home stereo. CD-ROM puts the storage capability of a mainframe system computer within your existing PC. This technology allows the storage of text, graphics, audio, video, video still frame, and animation—all in a digital form on a single CD-ROM disc. Today, commercially available, low-cost CD-ROM drives are easily interfaced with a PC to provide a cost-effective delivery platform for CD-ROM.

1.1 General Uses/Applications

CD-ROM technology has the flexibility for use in a variety of applications, including storage of technical manuals and archival data for quick reference, retrieval, space savings, and distribution; storage of bibliographic and on-line databases to assist in rapid search and retrieval of data through cross-referencing and indexing; and multimedia applications and interactive training.

1.1.1 Storing and Distributing Data

CD-ROM is ideal for storing large volumes of information that need to be distributed to many people in many locations. This form of information distribution is very cost effective when compared to the expense of printing, copying, and distributing the same information in paper form. If information needs to be updated periodically and has wide distribution requirements, CD-ROM can be a cost-effective method.

1.1.2 Research Databases

CD-ROM can replace on-line, bibliographic, and card catalog databases while providing quick access, cross-referencing, and retrieval of information. CD-ROM can provide access to these databases without regard to online connect charges, location, or time of day.

1.1.3 Multimedia Applications and Interactive Training

CD-ROM is ideally suited for the development of multimedia courseware and instructional/orientation programs. Programs can be developed on CD-ROM to create interactive exhibits, seminars, educational programs, demonstrations, training programs, and orientation packages. The multimedia aspect of CD-ROM (incorporation of text, graphics, audio, and animation) can create interest, attract users to the program, and stimulate learning. CD-ROM is best suited for distributing large amounts of courseware.

1.2 Capabilities/Advantages

As an emerging technology, CD-ROM has many capabilities and many advantages over other media, like paper, for storing and readily accessing large volumes of data.

1.2.1 Flexibility

A CD-ROM disc can store many types of data, including text, graphics, animation, audio, digitized photos, digitized video still frames, full motion video, and computer programs (software).

1.2.2 Storage Capabilities

A single CD-ROM disc can hold up to 650 megabytes of information. This is equal to about 325,000 pages of written text or more than 460 computer floppy disks (See Table 1 and Figure 1). A single CD-ROM disc can hold 6,000 graphic images, up to 72 minutes of stereo-quality audio, or up to 72 minutes of full motion video. It is possible to mix data formats on a single disc; for instance, 35 minutes of audio can be stored with 162,000 pages of text.

Medium	Capacity
Paper (ascii text @ 2 kb/page)	325,000 pages
Microfiche (98 frames)	3,316 fiche
Floppy Disk (1.4 mb)	460 disks
CD-ROM (650 mb)	1 disc

Table 1. Storage Capacity Comparison

CD-ROM is perfect for managing, archiving, and accessing data that are occupying valuable storage space. A single CD-ROM disc can hold data equivalent to 90 linear feet of shelf space and weighing 5 tons. This means that several filing cabinets worth of information or an entire encyclopedia set can be stored on one CD-ROM disc. Because of this capability, CD-ROM is ideally suited to applications where storage space is limited. The size of the CD-ROM disc also allows proprietary and classified information to be easily secured in a locked desk.

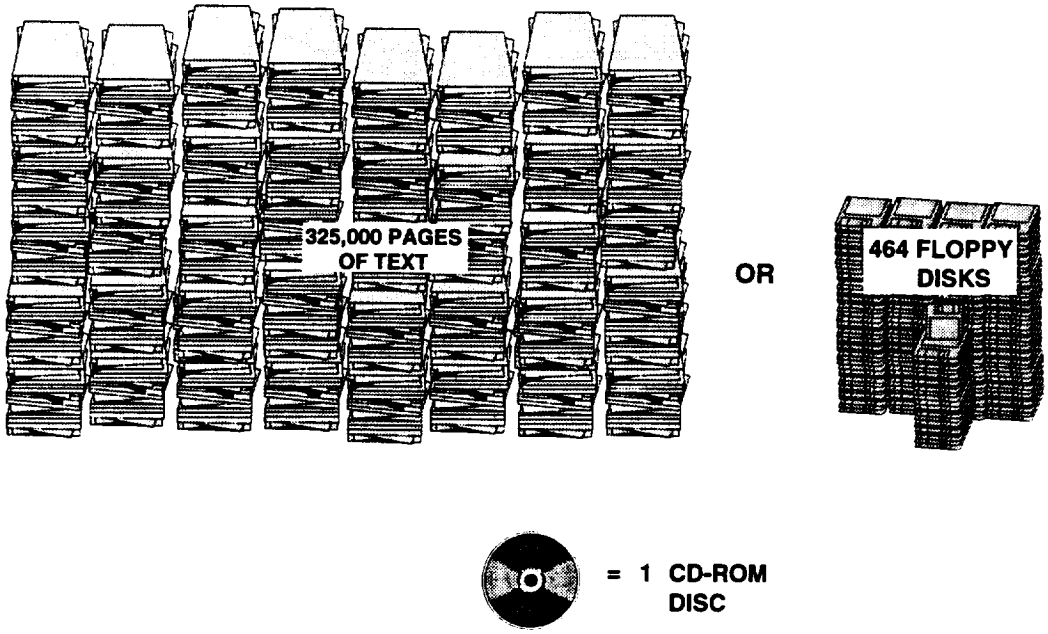


Figure 1. Storage equivalents; paper or floppies vs CD-ROM

A concrete example from the Government Printing Office compares the publication of the complete United States Code in paper versus CD-ROM (see Table 2).

Paper	CD-ROM
30,000 pages	24-page User Manual
24 bound volumes	1 CD-ROM disc
5 feet of shelf space	1/2" of shelf space
150 pounds	Less than 1 pound
$1,235 per set	$34 per set
Manual Search	Full Text Searching
Basic and Supplements	Single Database

Table 2. Comparison of paper-based storage versus CD-ROM.

A final comparison of the volume of information that can be stored on a CD-ROM and the savings that are possible with the technology comes from the United States Geologic Survey (USGS). Five years ago, an oil company wanted a copy of all the ocean floor mapping data and charts compiled by USGS. At the time, the information was collected on 600 9-track computer tapes and sold to the oil company for $80,000. USGS realized that the data was vulnerable because there was only one set of tapes, and they were extremely costly to

reproduce. CD-ROM technology enabled USGS to move the data onto 39 CD-ROM's and make the information available for $475.

1.2.3 Costs

While the initial production costs of CD-ROM are significant, once the disc is mastered, duplicate discs may be produced at a very low cost. Duplicate discs are needed for distribution or updating to keep information current. Duplicate CD-ROM discs cost from $1-$2.50 depending on the vendor chosen and the quantity duplicated (i.e., $1 per disc for 10,000 duplicates or $2.50 per disc for 500 duplicates). The costs for storing information on CD-ROM are also very low. For comparison purposes, the following are storage costs per megabyte of data: CD-ROM, $0.024; floppy disk, $0.35; hard disk, $0.52; microfiche, $0.76; and paper, $4.00 (See Figure 2).

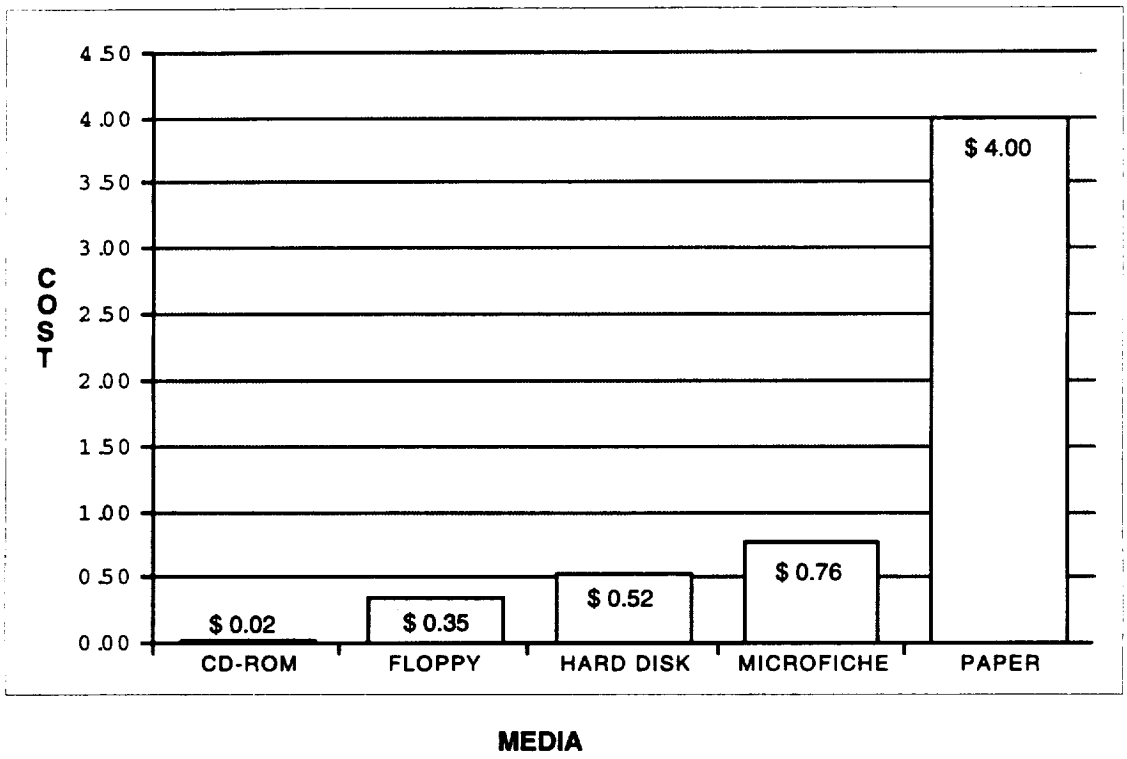


Figure 2. Estimated storage costs of different media (per megabyte of data)

1.2.4 Distribution of Information

CD-ROM is an excellent medium for distributing vast quantities of information to a large population of users in many locations at a very low cost (see Figure 3). It is extremely expensive to print, copy, and distribute large quantities of paper information. For example, 325,000 pages of text can be stored on a single CD-ROM disc. Each duplicate disc can cost as little as $1. The cost of copying this data in paper form, at an average of $.015 per page, would be $4,875 per set. To mail 325,000 pages of text would cost approximately $2,700 in

fourth class postage. Because the CD-ROM disc is small, it can be distributed to end users quickly and at a low cost (a CD-ROM disc in its plastic case can be sent as first class mail for $1.24).

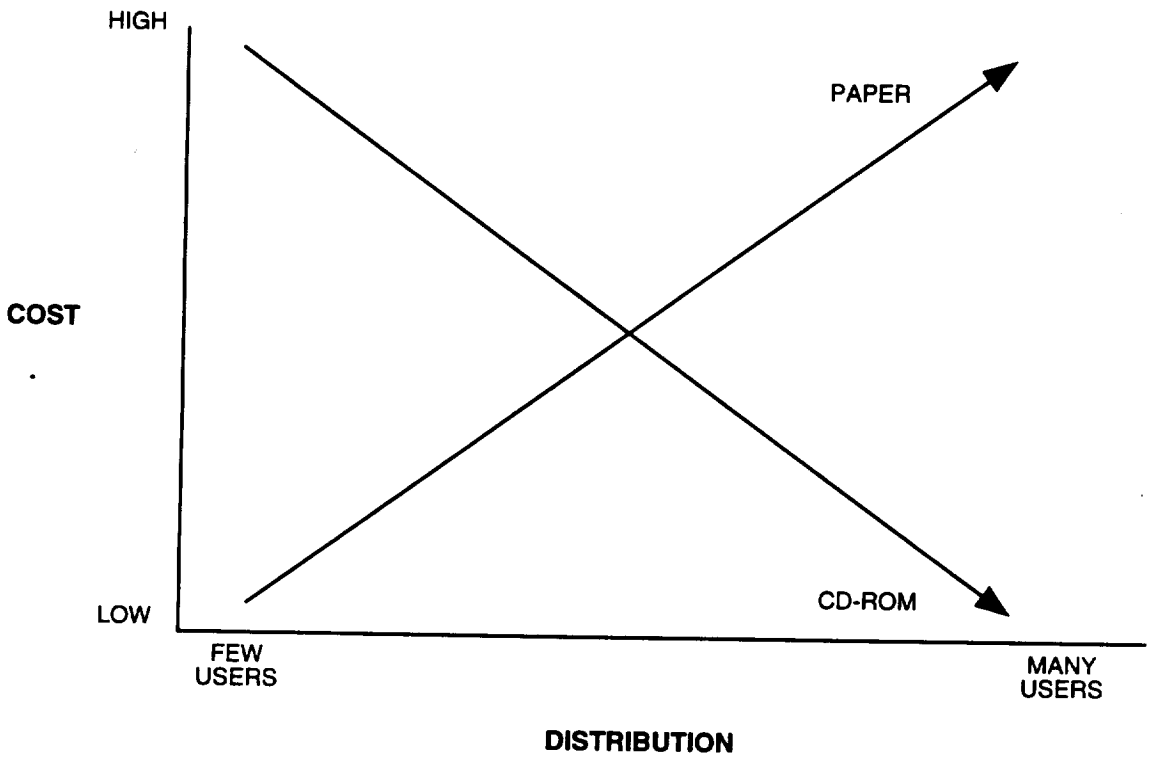


Figure 3. Distibution costs (paper versus CD-ROM)

1.2.5 Ease of Retrieval

Unlike a book, information stored on a CD-ROM is easy to access because of the high speed of the computer and the method developed for data retrieval. Information, not just text, can be retrieved from a CD-ROM disc. If a single fact has to be retrieved, it may be as easy and quick to look it up in a book as it is to search a CD-ROM disc (See Figure 4). However, if many facts have to be retrieved and information must be researched and collated, CD-ROM provides a clear advantage over paper. Unlike its paper counterpart, information on CD-ROM can be cross-referenced easily because it is possible to create hypertext links and perform keyword searches of data. CD-ROM can be indexed to search data by subject, title, keyword, or other descriptors, depending on the capabilities of the indexing software. Data stored on CD-ROM can also be readily printed or downloaded to a computer's memory for research, writing, and editing purposes.

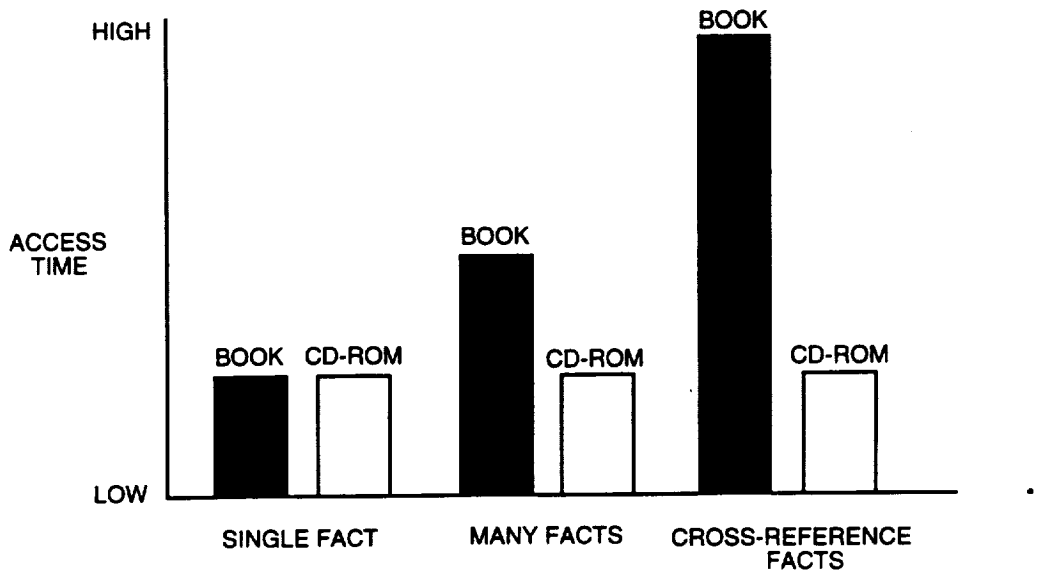


Figure 4. Retrieval Capacity of CD-ROM

1.2.6 Data Permanence

The CD-ROM disc itself is physically durable and can withstand extremes in environmental conditions. Data on a CD-ROM disc is highly reliable (resistant to damage), unlike any type of magnetic storage device because CD-ROM discs are not affected by dust or minor surface scratches. And because a CD-ROM disc is read by a laser beam, the optical head mechanism does not touch the disc. Therefore, no matter how many times the CD-ROM disc is read, it cannot be worn out or damaged. Unlike other computer disks, in the event of a computer failure, there is no chance of damage to the CD-ROM disc. Further, the shelf life of a CD-ROM disc is estimated to be 10 to 100 years.

1.2.7 Standardization

CD-ROM discs and drives are standardized worldwide. This means that CD-ROM discs are compatible with all kinds of computer systems, including PC's and mainframes. The structure of CD-ROM data files is governed by an international standard, ISO 9660, so that all CD-ROM data files can be structured to be read by any CD-ROM drive currently manufactured.

1.2.8 Commercial Availablity

CD-ROM drives are commercially available at a relatively low cost as are a large number (thousands) of off-the-shelf titles covering many subject areas.

1.2.9 Security

The ability of CD-ROM to store information depends on the physical integrity of the disc. If the disc is broken, the information on it is unrecoverable. In addition, ASCII or other data stored on the disc can be encrypted by any of several methods, with the decryption key stored elsewhere. These characteristics may make CD-ROM an appropriate storage medium for sensitive information. However, the small size of the disc makes it easy to hide, steal, or lose. These factors as well as the time, cost, and specialized equipment required to produce master CD-ROM discs may limit their usefulness in storing classified information.

1.2.10 Read-Only Memory

Prepared CD-ROM discs can only be read; users cannot edit or alter the data on the disc like they can with magnetic media. This provides an obvious advantage for data, like books and reference materials that should not be manipulated. Further, the inability to alter data coupled with long shelf life makes CD-ROM discs a great medium for archival information storage.

1.3 Limitations/Disadvantages

1.3.1 Read-Only Memory

Except for CD-R (CD-Recordable) discs discussed below, CD-ROM's are read-only sources; the information on the disc cannot be altered edited without creating a new disc.

1.3.2 Slow Access Time

The average access time for CD-ROM, about 300 milliseconds, is slow in comparison with the average access time of a computer hard drive, which is around 12-14 milliseconds. However, there are indexing techniques that can be employed to improve data access time.

1.3.3 Slow Data Transfer

While newer, higher speed CD-ROM drives are available and under development, the data transfer rate of commonly available CD-ROM is 150k bytes per second. Again, this is considerably slower than most hard drives or the computing speed of most current desktop computers. The higher the transfer rate, the better the performance and the smoother the playback of video and animation.

1.3.4 Production

The process for producing and publishing a CD-ROM disc can be lengthy and involves the use of vendors, technical experts, and/or specialized equipment. For instance, the organization of the indexing and retrieval system required for ready access to the information is a time-consuming process in itself.

1.3.5 Cost

The initial production costs of CD-ROM can be very high. The costs descibed below are physical costs and do not include the cost of equipment, preparation software, or labor. The commercial production cost of a CD-ROM disc can be prohibitive if little data needs to be stored or if the data need only reside in one location. The cost of creating a master CD-ROM disc is between $900 and $2,000, while the cost of duplicate CD-ROM discs are about $1 to $2. All duplicates are created from the master; if a sufficient number of discs are duplicated, CD-ROM is very cost effective. For single copies and small runs, CD-R (CD-Recordable) offers great promise at about $15 per disc.

1.3.6 Licensing

A licensing agreement with a vendor is necessary when using a vendor's indexing and retrieval software to access the data. Depending on content, other licensing agreements and royalty fees may be necessary.

1.3.7 Data Display

Because one of the advantages of storing information on CD-ROM is to be able to readily access the information, it must be stored in a text format (ASCII) that the computer will understand. This means that the information may not look the same as it does in the original document. For scanned documents, the spacing layout, character fonts, type size, and page breaks may be different, depending on the Optical Character Recognition (OCR) software utilized. For instance, fancy fonts can disappear if the OCR software does not recognize them. It is possible to have the information stored on CD-ROM as an image so that it looks exactly like it does in the original document (*WYSIWYG—What You See Is What You Get*), but then the information would not be retrievable and would instead be an electronic page-turner, using considerably more memory space on the disc.

1.4 CD-ROM Technology

1.4.1 Media

A CD-ROM is an optical, digital data storage medium. Optical storage devices are "read" by lasers through a method analogous to radar. Fluctuations in the laser's light caused by the imprinting of information on the CD are interpreted digitally. By contrast, floppy disks are magnetic storage devices; distinctly aligned magnetic particles are passed over by a "read" head that contains a wire. The magnetic particles start an electric current in the wire that then vibrates; the frequency of the vibrations is interpreted as data.

CD-ROM technology for information storage evolved from music compact discs. The disc is made from polycarbonate plastic with additional layers of aluminum, lacquer and paper (label). It is about 4.75" (121 mm) in diameter and about 1.2 millimeters thick (.047") (See Attachment 1).

1.4.2 Data Encoding

The polycarbonate layer contains "lands" and "pits" that represent the binary coding that forms bytes that form letters, words and images (See Attachment 2). Digital data are represented by combinations of "0's" and "1's." These two digits are the parts of the binary number system that are the foundation of computer technology. A "0" represents "off" and a "1" represents "on." Digital signals are either on or off with no middle ground. The "lands" and "pits" are read by a laser that interprets the variations in light as 0's or 1's. When the laser reads a land, a 1 is registered; when the laser hits a pit, a 0 is recorded.

1.4.3 Physical Layout

The disc is divided into tracks and sectors. A track contains a single spiral pattern of pits and lands. Up to 99 separate tracks can fit on one CD, with a total of 20,000 spirals from center to outer edge.

1.4.4 Reading a CD-ROM Disc

CD's, whether audio or data, are read from the center outward. The disc spins at a Constant Linear Velocity (CLV); the inner tracks spin at 550 rpm while the outer tracks spin at 200 rpm. This is an important issue because it affects the time it takes to locate information on the disc. Moreover, this difference in spin is the reason that premastering software is very important—it allows for data to be placed in the optimal position on the CD. Information that will be used or searched more frequently should be placed close to the beginning of the disc where the laser has less distance to travel. Less frequently needed information can be placed on the outer tracks. In order to interpret the data correctly and to locate the proper track, a sector at the beginning of the disc contains information about the synchronization coding, data location information, and information used to detect and correct errors.

1.4.5 CD-R

CD-R stands for Compact Disc-Recordable, a method of preparing individual or "one-off" CD's. CD-R's differ from CD-ROM in media—the laser in the CD-R device records information by acting upon an added layer of a special dye. CD-R technology allows the benefits of CD-ROM technology without associated mastering costs and turnaround time used for mass production CD's. CD-R's are important as archival and limited distribution mechanisms and are used as one-offs for custom databases. CD-R's also allow for beta-testing of a CD. Comments from users can be incorporated before the final CD is mastered and distributed. Using "authoring" and "premastering" software, the medium is also ideal as a method of transmitting the data to a mastering house for use in preparing large numbers of CD's. As systems evolve for multisessioning (the ability to record information at different times, on different tracks), the use of CD-R for daily, weekly, and monthly archiving of data will also grow. Finally, these systems are stable and unaffected by electrical or magnetic pulses: minimum estimated shelf life is ten to twenty years, with special discs guaranteed in excess of 100 years.

1.5 CD-ROM Standards: An Introduction

Standards are defined rules accepted and followed by a group having similar needs; standards allow for a common method of development to take place. Standards can be divided into voluntary and regulatory standards. Voluntary standards are de facto, market driven and unenforceable, whereas regulatory standards are generally more rigid, more clearly defined, and, having been issued by an entity with statutory power, more enforceable. Standards help provide continuity to protect systems from the erratic conditions of the open market and help enable equipment and systems from different manufacturers to work together without collaboration.

1.5.1 IEC 908 (Red Book)

The earliest and fundamental CD-ROM standard was established by Philips and Sony. The Compact Disc Digital Audio Standard, IEC 908, commonly known as the Red Book standard, coincided with the introduction of the audio CD. The Red Book standard specified both the structure of the CD-Digital-Audio (CD-DA) track and the mechanism for data error detection and correction. Tracks were defined as individually recorded selections (song, musical movement, etc.) and subdivided as fixed units of fixed length and duration. These units were termed sectors, having a duration of 1/75th of a second and containing 2,352 digital bytes of information.

1.5.2 ISO/IEC 10149 (Yellow Book)

Philips and Sony further defined standards with the issuance of ISO/IEC 10149, the Yellow Book. Based upon the Red Book, the Yellow Book standard defined two new track types: CD-ROM Mode 1, for computer data, and CD-ROM Mode 2, for compressed audio data and video/picture data. The Yellow Book made it possible to combine data files and audio information on a single CD. A CD with combined Mode 1 and Mode 2 tracks is referred to as a "mixed mode" disc. Mixed mode discs require separate tracks for data files and audio information. A mixed mode disc cannot read data files while it is playing audio.

1.5.3 CD-ROM/XA

Jointly, Philips, Microsoft, and Sony extended the Yellow Book standard and defined a new track standard known as CD-ROM/XA. The Yellow Book CD-ROM track definition was expanded, allowing the interleaving of data files and audio sectors on the same track. A CD-ROM/XA is capable of reading data files while playing an audio selection.

1.5.4 Green Book

The Green Book standard was developed for Compact Disk Interactive (CD-I). Principally, CD-I systems consist of stand-alone players connected to television sets. Green Book sector layouts are identical to CD-ROM/XA.

1.5.5 ISO 9660

ISO 9660 is an International Standards Organization (ISO) standard that describes the structure of computer files to be placed on a CD-ROM. It is an outgrowth and modification of a standard commonly called High Sierra. The High Sierra standard was an early industry attempt to format CD-ROM data without regard to operating system. Interfaces to ISO 9660 have been developed for many operating systems, including MS-DOS, Apple's HFS, and UNIX. The ISO 9660 standard specifies directory file structure and nomenclature. These structures were originally based upon the DOS operating system and have created additional concerns for Apple and UNIX uses. UNIX vendors (the Rockridge Group) are in the process of developing additional ISO 9660 standards to solve the UNIX limitations of the original standard.

1.5.6 Orange Book

The Orange Book provides standards for recordable CD. Part I, provides standards for CD-MO (Compact Disc–Magneto Optical), that can be written, read, and erased. Part II describes the standards for CD-WO (Compact Disc–Write Once). Both Parts I and II allow the recording and playback of audio, video, and computer data.

2 CD-ROM Production/Publishing

To take an original document (source data) and publish it on CD-ROM is an involved and time-consuming process consisting of several distinct steps that require expertise in areas of instructional design, data conversion, quality assurance, search structures, CD-ROM publishing, and, possibly, computer programming. (Figure 5 depicts the CD-ROM production process.)

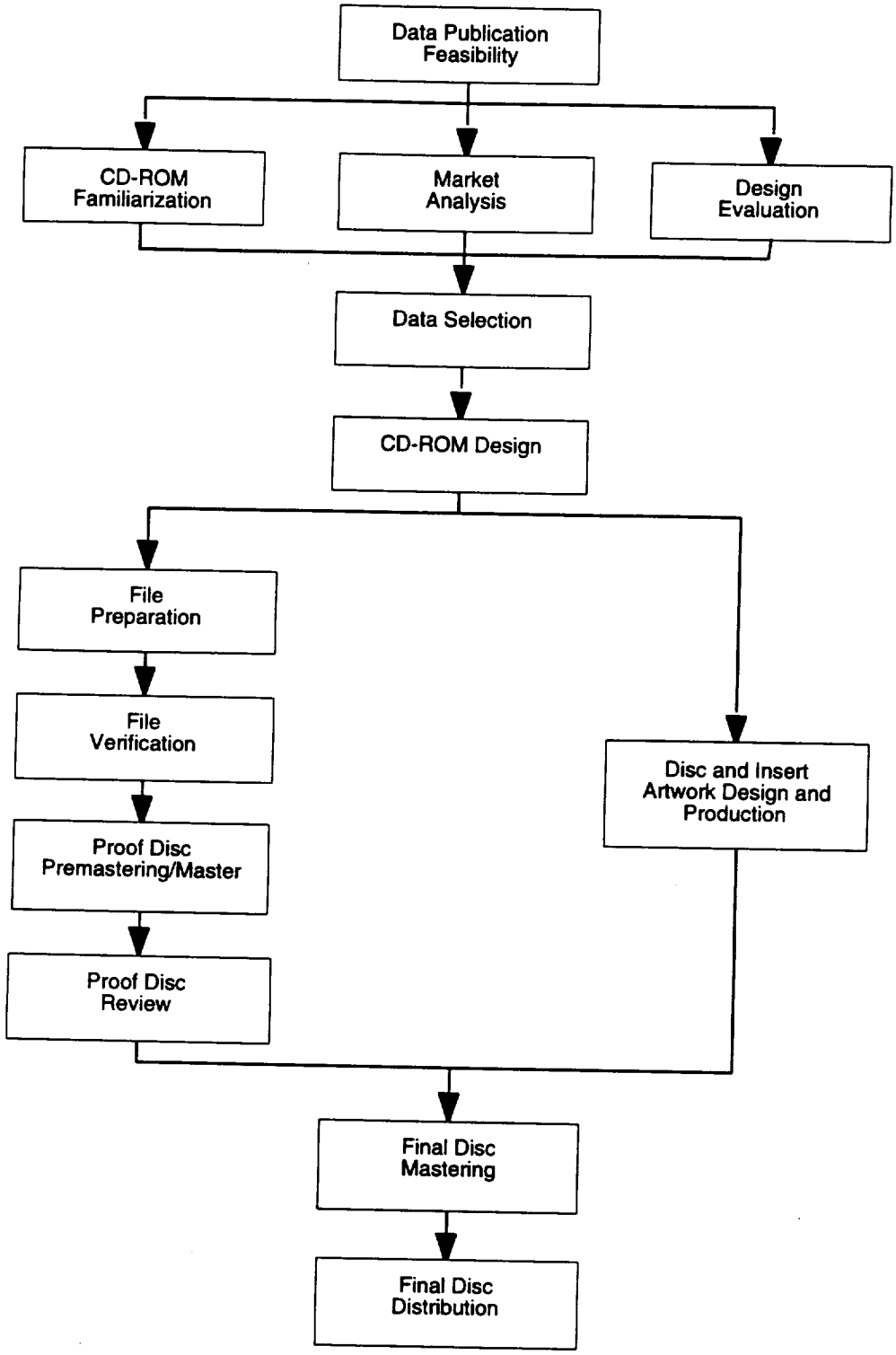


Figure 5. CD-ROM production processing flow

Prior to starting a CD-ROM project, two kinds of review should be made—a market analysis and a data analysis. Market analysis is beyond the scope of this project, but before producing a CD-ROM some questions have to be answered, since the answers will help determine the kind of software and documentation needed to start the project.

— Who will use the product? Who is the audience?
— How will the product be used? Why do they need it?
— How will the product be marketed? How will we reach the audience?
— What kind of features will be required? Are there special needs?
— What documentation will be needed—a lot or a little?
— What help should be available?
— What is likely demand?
— Will updates be required?
— What costs are associated with preparing and delivering the product?

Data organization and structure: The data analysis review consists of four parts (see Attachments 3 and 4).

— What kind of information will the data represent?
— How much alteration of the data will be needed to prepare it for a CD-ROM?
— How much data will be included?
— How will the data to be searched and retrieved?

Answers to the above questions will determine the kind of authoring software (see Attachment 5) that will be used to make the CD-ROM. It is important to keep the requirements of the authoring software in mind when preparing the data. How the finished data will be retrieved and viewed on screen is determined in large measure by the software requirements. The software configuration and completion of the time and cost estimates involved in developing the CD-ROM will depend upon combining the results of the market analysis and the data analysis.

2.1 Design

Before actual CD-ROM production begins, many factors must be considered so that each of the production stages is performed efficiently and leads to the desired result. Personnel with CD-ROM production knowledge and database design expertise should be consulted before undertaking a major CD-ROM database project. Some other considerations, such as source data, operating platform, and the user interface, are described below.

2.1.1 Source Data

Consideration must be given to the form of the data that will be published on CD-ROM. Is it paper documentation, microfilm, microfiche, on computer floppy disks, or on a computer's hard drive? What is the condition or quality of the source data?

2.1.2 Type of Information

The type of information—text, photographs, graphics, etc.—must also be determined. If the information consists of text, is it fielded data like a spreadsheet or a database, or is it textual information like a book or standard document? This information will help determine the computer memory and resolution requirements necessary for later production stages. Some resolutions are best for print, others for screen display.

2.1.3 Delivery Platform

Consideration must be given to the capabilities (storage capacity, graphics display capability, etc.) of the platforms (computer system) that will be used for accessing data on CD-ROM. It does no good to incorporate fancy features into CD-ROM if the user cannot take advantage of them because of limitations of the computer system.

2.1.4 User Interface

The structure of the data must be given careful thought so that an indexing and retrieval software package may be selected. This software is selected on the basis of its particular features for designing a database and the target hardware system configuration (delivery platform). The indexing and retrieval software is used to create a user interface that allows easy and intuitive use of the CD-ROM system by the end user. Creating a user interface involves designing the layout of the computer screen, menu, and function key structure. Indexing also involves designing the database so that information can be readily cross-referenced, searched, and retrieved. Data preparation, especially design of the indexing and retrieval system, can be the most difficult and most labor intensive part of the CD-ROM publishing process. The data design stage is the time to consider how data searches will be conducted and how the information will be retrieved (by keyword, as paragraphs, names, images, etc.). The data should also be structured in a standard way so that later revisions can be handled easily. (See Section 2.4.2 for more information on indexing and retrieval.)

2.2 Data Conversion to an Electronic Medium

If the print material is not in an electronic medium (i.e., computer files) it must be converted so that data manipulation can be performed later in the production process. Non-electronic source material includes paper documentation, microfiche/microfilm, photographs, graphics, maps, etc. Data on floppy disks or a computer's hard drive are already in electronic form.

Print material can be converted in three ways: keyboard data entry, image scanning and OCR scanning. Data conversion, whether it be for paper documents, microfiche or microfilm, or images, can be done in-house with the appropriate equipment (PC, Optical Character Recognition (OCR) software, and scanner) or it may be sent to a vendor who will perform these services.

Text is converted from raw data to formatted data—data that has been "tagged" or uniquely identified in order to allow the "build" software to develop a structure enabling the user to retrieve specific bits of information. Tags are used to indicate new chapters, sub-chapters,

paragraphs, phrases, keywords or any other element that is to be made searchable. One example of a very structured tagging system is Standard Generalized Markup Language (SGML). SGML allows full-text tagging in a widely accepted format. IBM, the American Association of Publishers, and others use SGML for their documents.

2.2.1 Paper Documentation

Data in paper format can be converted by either entering the data onto a PC via keystroking (keyboarding) or by electronically scanning the document and processing it through OCR software. Scanning a document stores it as an image on a magnetic medium. An image cannot be manipulated like text can in a word processor. In order to manipulate text that is in an image form, the image must be converted to a format that is machine-readable. Running the scanned text image through OCR software converts the data from an image to a collection of its searchable components (i.e., individual words). Both keystroking and OCR scanning are labor intensive.

2.2.2 Microfiche/Microfilm

The conversion of microfiche/microfilm materials to an electronic medium can be a problem. Although many companies claim the ability to convert microfiche directly to ASCII text by scanning, the results are poor and in many cases may not be usable. With advances in the application of CD-ROM technology being made every day, it is possible that in the future this may be a viable option, but the technology is not quite there. If microfiche/microfilm data are desired on CD-ROM, it is recommended that the data be obtained in its original form (as a paper copy or, possibly, magnetic medium) and converted to an electronic form. If a great deal of microfiche or microfilm data will be stored on CD-ROM and produced over a period of years, consideration should be given to converting data from a paper form to an electronic medium before or in place of putting it onto microfiche/microfilm.

2.2.3 Photographs and Graphic Images

Photographs and graphics can be converted by either an electronic scanner or by using a video camera. There are advantages and disadvantages with both methods which must be considered before a choice is made. Use of a video camera is probably best where volume is great. Scanning graphic images and photographs is more labor intensive than scanning text documents because the resolution (number of dots per inch) changes, frequently requiring an artist to touch up jagged edges on artwork or redraw some lines. Graphics and photographs are not processed through OCR software.

2.3 Quality Assurance

The machine readable converted version of the document should be checked for accuracy. This involves comparing the original source document to the computer file. Again, this is a time consuming process depending on how much data needs to be processed and how many errors in typing or scanning occurred. Scanning good quality documents using OCR software typically has a 95-99% accuracy rate. For a typical page of text, this means that there could be between 20 and 100 errors per page. Correction of these errors will require a substantial

word processing effort. For instance, source documents with different fonts and type sizes in the same document can produce many errors when using OCR. Entering the data on a PC by a qualified typist may produce fewer errors, depending on the sophistication of the OCR software. OCR software packages vary; some incorporate a verification stage that reduces the amount of spell checking and quality assurance required. Today, there are high-end OCR softwares that provide rapid document conversion and produce very few errors. The trade-offs between creating a data file via keystroke or OCR scanning and the editorial time needed to correct errors must be considered.

Quality assurance can be done in-house with the appropriate equipment (PC and word processing software) or this may be done by a data conversion service vendor. However, not all vendors will perform quality assurance of a document after it is scanned.

2.4 Data File Preparation

In order to take advantage of the search and retrieval capabilities of CD-ROM technology, the source data must be indexed. Indexing the data is the process of setting up the data files so that the data can be retrieved. The sophistication of the indexing and retrieval system will determine how user friendly the system is and how accessible the data are to the user. Indexing is one of the most critical steps in developing a CD-ROM product, for it can enable the user to access information in ways that facilitate research, stimulate learning, and facilitate information transfer. Simply, it can put the user in control and let him decide what is significant and what is not.

2.4.1 Index Creation

The first step in indexing is the creation of the user interface. This is done by designing the layout of the computer screen and the menu and function key structures. The second step in indexing involves structuring the data to a recognized standard governing the type of data (e.g., text, graphics, audio, video). The data format employed must be compatible with the indexing and retrieval software used to structure the database for ready access. Without this organization, the CD-ROM disc becomes nothing more than an electronic page-turner. The retrieval method is designed in accordance with the menu and function key structure. An experienced designer should perform both steps to ensure that the indexing and retrieval system is user friendly, efficient, and complies with standards.

"Authoring" software is the technical name given to the software that builds the indexes and establishes the linkages for the retrieval software. The "authoring" software contains the "build" engine that converts and indexes the data and the "retrieval" or "search" engine that allows the data to be found on the disc.

As mentioned in Section 2.2, the text is converted from raw data to formatted data—data that has been "tagged" or uniquely identified in order to allow the "build" software to develop a structure enabling the user to retrieve specific bits of information. Tags are used to indicate new chapters, sub-chapters, paragraphs, phrases, keywords or any other element that is to be made searchable. One example of a very structured tagging system is Standard Generalized Markup Language (SGML). SGML is a set of rules for defining generalized markup

applications. A markup language identifies text or sections of text and specifies what processing functions should be performed on them, for instance, bolding text, setting tabs, identifying section headings, or paragraphs. Markup allows data to be stored, searched, and accessed because it specifies information the computer needs to perform these functions. SGML creates an environment where the commands for tab sets, indents, justified margins, and hyperlinks, for example, are the same whether a document is a government publication, military publication, or newspaper. This enables easy transfer of data between government agencies and the private sector. The SGML standard is universal: a text file using SGML can be read by any commercially produced hardware or software. IBM, the American Association of Publishers, and others use SGML for their documents.

Since the idea of using a CD-ROM is based upon having ready access to large amounts of data, a search strategy that allows the user to search quickly and efficiently is vital. The "build" engine of the "authoring" software organizes data by indexing searchable words or tags before premastering. The indexes generated are "inverted indexes" (i.e., indexes built on the numeric frequency of each word, except stop words like *to, the, of,* and *an*), and placed in alphabetical order. The indexes also have pointers to the location of each word. Once the engine has built the indexes, it is no longer required and is not included with the CD-ROM.

The "search" engine portion of the software is written to the CD-ROM. It receives the requests for data and searches the CD-ROM's indexes and text for the desired information.

2.4.2 Text and Record Formatting

Full-text indexing lists every searchable word and phrase except "stop" words like *a, of, an, the*, and *to*. The indexes contain every word in alphabetical order. One might consider a novel, if it had an index, an example of a full-text database.

Fielded or record data are records of a fixed length. The data records in the fields are always in the same order from record to record. Names, addresses, zip codes, and phone numbers are examples of data within a fielded database. Search times are much faster with this type of database, and searching can be more powerful, since specific types of information can be searched for in specific places. A telephone book is an example of a fielded database.

With either the text or record format, the files (in the correct format), any compressed images, the indexes, the retrieval software, and auxiliary files are placed by the "authoring software" in the order that they will appear on the production disc and with careful consideration for the way that they will be used. Placing the files properly and in the order that a user will search them is critical to optimizing the speed of data retrieval from the disc.

Indexing full-text information requires less work than does fielded information. Data retrieval is performed through free-text searching using a variety of search types set up by the designer of the database:

- Keyword and subject searches allow the user to enter in a keyword for the computer to search. Generally, the number of occurrences of that keyword is displayed along

with their location in the text. The user simply clicks on the occurrence that he would like to view. A comprehensive word index in which every word in the text (except stopwords) appears is sometimes used as an index that enables the user to select the word he wants to search.

- Hypertext searches define associative links between data that, when clicked on, provide nonlinear viewing of information. This allows the user to follow his train of thought and examine information in any order he wishes and at any level of detail desired by selecting the highlighted topic he wants to view.

- Boolean searches (named after the mathematician George Boole) allow the user to combine terms with *and, or,* and *not* to refine or expand searches. For instance, if research is being conducted on remote sensing of Mars, the subjects *remote sensing* and *Mars* can be combined in one search. That is, both terms would be simultaneously searched instead of searching for all general references to remote sensing and then narrowing those to occurrences to Mars. Where a database is large, Boolean searches can save much research time.

All indexing systems take up additional storage space on the CD-ROM disc, in fact some can take as much as 1/3 of the disc's space. Therefore, it is not advantageous to have all words indexed for searches. Words that will not help a search, like *the, and, at,* can be eliminated from the indexing system. Indexing software packages generally come with a list of these "stopwords" that are excluded from the indexing system. This list can be modified to add or delete words to suit your needs. All words not on the stopword list will be indexed and be searchable.

There are many commercially available indexing and retrieval softwares packages that enable the user to create the data file structure. These packages should be reviewed carefully to assure compatibility with the selected database design and search and retrieval requirements.

Indexing can be done in-house with the appropriate equipment and personnel or this process can be done completely by an outside vendor.

2.4.3 Transfer of Data to Magnetic Medium

After indexing, data must be transferred to a portable magnetic medium so that it can be mastered onto a CD-ROM disc. Magnetic media include computer floppy disks, removeable computer hard drives, 9-track tapes, data cartridge tapes, and CD-R's. The type of magnetic medium chosen must be an acceptable input medium for the CD-ROM vendor who will master the disc. Data transfer can be handled in-house with the appropriate equipment or it can be contracted to a vendor.

2.4.4 Premastering the Data

Premastering software organizes the database into an industry specified format, ISO 9660, for logical formatting and permits placing the data, address blocks, and error correction

information in the optimal configuration on the disc. ISO 9660 is the accepted standard in the CD-ROM industry, allowing CD-ROM discs using this standard to be read by any CD-ROM drive currently manufactured. Most commercially-produced indexing and retrieval software also require CD-ROM discs to be in ISO 9660 format. Premastering is a one-step process that is controlled and completed by a computer. Premastering can be done in-house with the appropriate equipment (PC and a CD-ROM Premastering software package) or by an outside vendor. Premastered data can be sent to a mastering house to create master CD's or used to create one-offs using a CD-R.

2.4.5 Checking the Search Structure for Proper Retrieval of Data (CD-ROM Simulation)

After data premastering, it is important to check the search structure to ensure it does retrieve the data. This is a quality assurance process to validate the menu, function key, and data structure prior to mass production of CD-ROM discs. This can be done in-house with the appropriate equipment (PC and CD-ROM software with simulation feature) or by a CD-ROM vendor.

If it is done by an outside vendor, the following will occur:

- If the data is not premastered, the vendor will complete this process.

- The vendor will create one CD-ROM disc (master), which is called a check disc. The check disc is sent to the client along with the input data. The client will test the CD-ROM disc to validate the menu, function key, and data structure prior to giving the vendor permission to create duplicates. If any mistakes are found, the client will make corrections to the input data and return it to the vendor, who will create a new master.

  The check disc is an option that the client may or may not choose. There is usually an additional cost if a check disc is requested. The cost of the check disc is about a quarter of the cost of creating a master. If a check disc is not requested and the master disc is incorrect, the client will have to pay the full price for a new master to be created.

2.4.6 Mastering the CD-ROM Disc

The final step in CD-ROM production/publishing is the creation of the master disc and duplication of CD-ROM discs. Mastering is a four-step process that begins with the etching of digital data on a glass disc covered with a light-sensitive coating. Once a glass master is produced, a nickel "father" is made, followed by a nickel "mother" that is used to produce plastic "sons" that are used to press the consumable compact discs. Because this process requires specialized, costly equipment, it can only be performed by a vendor.

2.5 One-offs versus Volume Production

One-off production is enabled by the development of CD-R technology. CD-R's offer the opportunity to prepare and test information via authoring software and provide the complete

complement of features that would be available from a mastered disc but at lower cost and with quicker turnaround. The process works well for limited runs. Production time on a CD-R machine varies, depending on the speed of the drive. If a full disc of information were to be produced on a single-speed machine, it would take about an hour to make one disc; a 4X machine would produce a disc every fifteen minutes. Moreover, the cost of the CD-R discs is higher than that of discs prepared in quantity by a mastering lab. If the need for a particular CD was great enough, a single CD-R could be sent for duplication at a mastering facility and prepared economically and quickly, stamping a CD-ROM every five seconds. The question is quantity and mass production versus customization. Some of both will likely be required, and the capability of producing a one-off has additional benefits, such as the testing of retrieval software, data content, and organization.

2.6 Packaging

Packaging a CD-ROM can range from a simple cardboard container to elaborate jewel cases with text and images in four-color. The packaging should be determined by the need. In most cases the packaging will contain loading instructions on the CD and on the packaging. More elaborate packaging can be made up if it is determined through the market analysis that it is needed.

2.7 Documentation

Documentation is an extremely critical issue. While no documentation is an option, it is clearly not a good one. Mislaying the installation instructions is easy, rendering a disc unusable. Information about searching techniques or help screens, configuration requirements, and copyright information can also be misplaced or lost before the CD-ROM can be used. Preparing and printing the documentation may take as long as the preparation of the CD-ROM itself, so care and forethought will be necessary. The best method is to prepare the documentation that seems appropriate and then distribute a one-off CD-R to a novice user and ask for feedback that will be used to modify the final version. Documentation should be included on both the paper copy and the disc; this is a low-cost, simple way of insuring that users never lose the basic documentation.

2.8 Production Costs and Turn-around Time

Costs and time to prepare a CD-ROM will vary depending on the variations in requirements for data preparation, documentation, retrieval software licensing, quantities and whether the procurement is through GPO or a commercial vendor. GPO currently offers a set rate once the premastering tape is complete: $850 to prepare the master; $1.63/disc for 1,000 copies. The disc includes a silk-screened label in up to three colors, a jewel case and printed insert card. More elaborate packaging will increase costs slightly. Given the volume the STI Office prepares for the custom and continuing bibliographies, in the range of 100-250 copies, the costs per unit will be higher. Attached is a local vendor's current price schedule for comparison with the above GPO costs (see Attachment 8). The cost for mastering is less, but the costs per CD are higher.

Time to prepare the material will also vary. Most of the time will be taken up in preparation of the data, including the development of the documentation. Documentation should be included on the CD-ROM itself, at least to the extent of placing the loading instructions on the label. The documentation may include things like user manuals and instructions. Testing on a novice user can provide valuable feedback about the utility and usability of the disc and the instructions. When the data is in the correct format, the documentation has been prepared and tested, and the CD-ROM has been premastered, the final product can be produced within two weeks. Rush service, with a five-day or less delivery turnaround, can also be procured at additional cost.

3 CD-ROM Cookbook for CD-ROM Production

The following sections offer guidance in the stepwise production of a CD-ROM. While most steps are sequential in nature, just like baking a cake, it is possible to execute parallel instructions (*viz.* preheat the oven while mixing ingredients). Figure 6 below provides the major steps in producing a CD-ROM. As with most complicated processes, many decision points, branches, and possible loop procedures exist. The production of small CD-ROM quantities can be easily accomplished, at an affordable price, using 486 or Pentium® PC equipment with an attached CD-R (CD-Recorder), a one gigabyte hard drive, and CD-ROM premastering software.

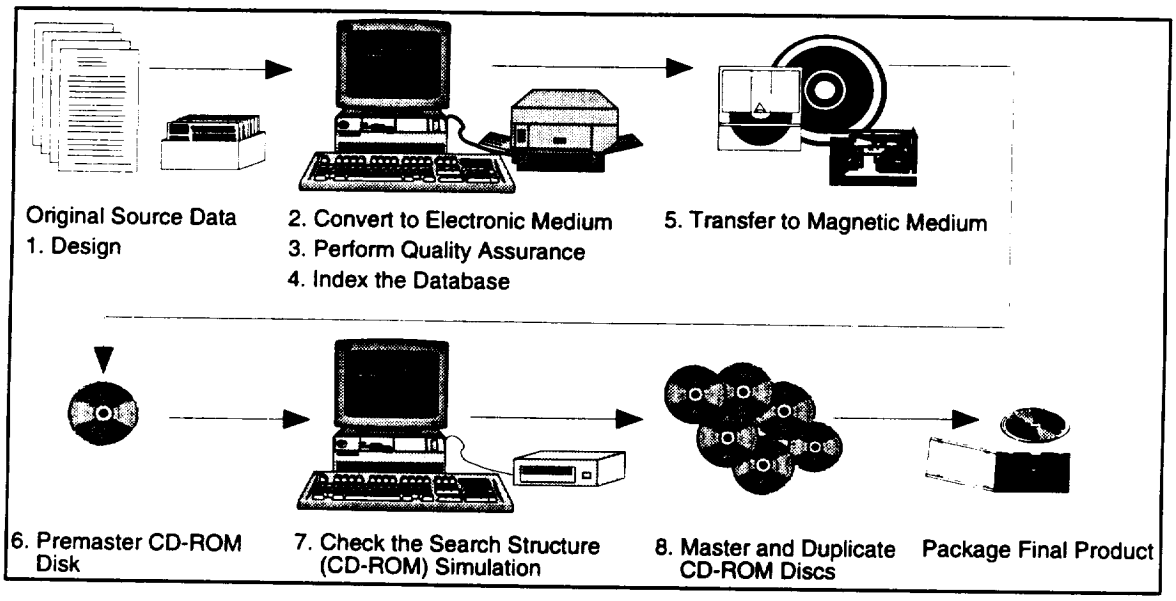


Figure 6. Stepwise CD-ROM Production

3.1 Use of CD-ROM Vendors

CD-ROM vendors can be grouped into three broad categories: publishing/production vendors perform all or specific steps required for CD-ROM production; equipment vendors sell the equipment necessary to produce CD-ROM or to read CD-ROM; titles vendors sell commercial off-the-shelf CD-ROM's that contain reference materials, educational packages, various periodicals, etc.

3.1.1 Considerations for Choosing a Vendor

When choosing a vendor for any one of the steps of CD-ROM production, for equipment or software purchases for reading CD-ROM, or for the purchase of commercial titles, careful planning and thought must be given to decisions. Vendors provide a variety of services, have varying levels of capability and stability, and charge a wide range of prices. Equipment and software can be purchased through private sector vendors that have established government contracts. Vendors, newspaper and magazine articles, and advertisements are a good source for some initial research. These sources of information

provide an overview of the prices, services, equipment, and software available. Visits to local CD-ROM dealers, chain stores, computer stores, and discounters are another way to obtain some initial information. However, when refining the search and selecting a CD-ROM vendor, the following factors should be considered.

- **Reliability.** With new applications for CD-ROM technology being developed every day, CD-ROM's are becoming more popular and the market is expanding, with vendors selling a variety of equipment, software, and services. All of these vendors promise to deliver a reliable product or service but not all vendors will be able to do so. That is why it is critical to check a vendor's reliability and customer satisfaction status before entering into an agreement. For instance, does the company specialize in CD-ROM or is that an area recently acquired? How long has the company been in business, particularly in CD-ROM technology? Can the vendor produce the phone numbers and addresses of satisfied customers? Before selecting a vendor, these questions must be answered. References should be contacted to learn about the vendor's performance in terms of product quality, meeting the customer's needs, adherence to CD-ROM standards, and timeliness of delivery. The vendor's financial stability should also be checked. Does the vendor have the resources to deliver what it promises? Does it have sufficient capital to keep up with current technology? Many publishing vendors require a considerable up-front investment—as much as 50% down, 25% at acceptance of design, and 25% on delivery. Young or small companies may not be able to deliver an acceptable product or service or may lack sufficient financing and go out of business before completing a contract. Recouping an initial investment could be very difficult under those circumstances.

- **Ability to Demonstrate Products/Services.** When choosing a vendor, insist on demonstrations, especially of software programs. Most companies can provide sample floppy disks of their CD-ROM software. Demonstrations will help ensure that the program fulfills the vendor's promises and that it meets technical and usability requirements. Many companies will provide evaluation copies of software.

- **Customer Satisfaction.** Before selecting any vendor, it is important to ask for the names, telephone numbers and addresses of previous customers. These companies or individuals should be contacted to get information about the vendor. The following questions should be asked: Was the customer happy with the vendor's services or products? Was product quality good? Were delivery schedules met? Was a representative readily available for consultation when necessary? Would the customer use that vendor again? If possible, the product(s) the vendor supplied to the customer should be reviewed.

- **Licensing/Royalty Fees.** Almost all vendors charge a licensing fee for use of their indexing and retrieval software programs. Many also charge a royalty fee on the duplicate discs made from each title produced using that software. This applies to both use of the software in-house and the vendor's use of this software to produce titles. Some vendors claim to have no royalty fees and use terms such as annual "retrieval license" for per-disc charges. But whatever the term, most vendors charge some form of licensing/royalty fees for the use of their software. The only way to

avoid this cost is to program indexing/retrieval software in-house. Licensing/royalty fees are charged in various ways:

- A royalty fee is paid per disc produced. This fee can range from $5 - $130 per disc depending on the number of duplicates made (e.g., $5/disc for 5,000 duplicates or $130/disc for 100 duplicates).

- A flat rate is paid for a minimum of 100 duplicates or a higher flat rate for unlimited duplicates. This rate can range from $6,000 - $12,000.

- A flat rate is paid for each CD-ROM title produced using the software, regardless of the number of duplicates. This rate can range from $750 - $10,000. This fee may decrease for each additional title produced.

- A flat one-time fee is paid for indexing and retrieval software use with no limit on the number of titles and duplicates created. This fee is usually about $30,000 - $60,000. This fee covers the cost of the software and its licensing.

- **Costs.** CD-ROM is still a relatively new technology with new vendors entering the market constantly. In order to capture new customers, many of these companies will negotiate prices aggressively. Therefore, there is a wide range of prices for CD-ROM services and equipment. For instance, because CD-ROM drives are standardized, the buyer can shop for the best price. Although generally directed toward individuals rather than organizations, many companies now offer package deals that enable the purchase of a drive together with one or more currently available titles. This can be a very economical way to purchase drives and CD-ROM discs. However, keep in mind that the price of drives reflects performance. For instance, products offering rapid retrieval time and/or a faster transfer rate will generally cost more than drives with less desirable capabilities. Opt for drives with the best retrieval time and transfer rate.

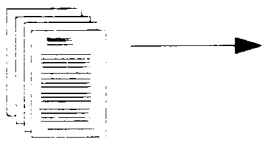
- **On-going Vendor Relationship**. When periodicals or other materials that are updated frequently are produced in the CD-ROM medium, an on-going relationship may be established with a vendor. That is, the vendor will produce the initial CD-ROM and at regular intervals (e.g., quarterly, semi-annually, annually) produce additional CD-ROM's to capture current editions or issues of the material. This on-going work and its frequency are factored into price negotiations and normally result in a lower cost for the purchaser than if he began negotiations anew each time an updated CD-ROM was to be produced.

3.2 CD-ROM Production

Production of CD-ROM is a time-consuming, complex process. It requires time, money, and personnel with expertise in instructional design, information retrieval, medium transfer, computer programming, etc. Decisions must be made about whether to work with a publishing vendor or to make the time, training, and equipment investments that enable in-house production of CD-ROM's. This section will present the considerations for

planning CD-ROM in-house, the equipment involved, and the use of vendors for each step of CD-ROM production.

3.2.1 Step 1 - Design



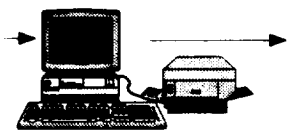
Original Source Data
1. Design

In-house Staff. To design a database, in-house staff must have expertise in database design so that they may select an appropriate indexing and retrieval software package (if one is not developed in-house) appropriate to the delivery platform chosen, the database, and the skill level of the targeted end users of the system. It is recommended that an indexing and retrieval software package be purchased from a vendor, particularly if many CD-ROM's will be published by the STI Office. Alternatively, other NASA Heaquarters Codes publishing CD-ROM's should be contacted for possible joint-licensing agreements. Experienced individuals should be consulted to assist in the design of a user friendly database and applications that make effective use of colors, screen layout, menu, online tutorials (optional), and function key structure.

Equipment. No equipment is required in the design phase.

Vendor. Many vendors provide consulting services concerning the CD-ROM design process.

3.2.2 Step 2 - Data Conversion to an Electronic Medium



2. Convert to Electronic Medium
3. Perform Quality Assurance
4. Index the Database

In-house Staff. Data conversion—via keyboard entry, image scanning, and/or OCR scanning—can be performed in-house with the appropriate equipment. However, some expertise in data conversion is required, and all three methods are labor intensive (although image scanning is more time consuming than text scanning). The selection of a sophisticated OCR software package for scanning text that produces few errors is important, especially if the work is done in-house. The scanning process will require skill in the use of specialized equipment and, if images are scanned, possibly graphic arts. Cost as well as the pros and cons of staff experience and available time should be weighed in making a determination for converting data in-house or using a vendor's services.

Equipment. A PC (486, 33 MHz or better with 8Mb RAM) with a large hard drive of at least 1 gigabyte of memory is necessary to store the database and enter the data via the keyboard. In addition, a scanner is needed for the transfer of source materials (graphics or text) to a magnetic medium. A flatbed scanner should be used; they can range in price from $500 - $6,000, depending on features and the amount of software included. Hand-held scanners should not be used because, with a scanner head width of 4", they are generally designed for graphics and clip art and not 8½" by 11" documents. However, hand-held scanners are useful for scanning archival documents that require special handling or rare books that cannot be taken apart so their pages can be placed flat on a flatbed scanner. Hand-held scanners range in price from $100 - $500, and their image conversion resolution is not as high as a flatbed scanner.

If text is being transferred, OCR software is needed. OCR software can range in price from $300 - $1,000, depending on its sophistication for recognizing characters and producing fewer errors. Consideration should be given to these higher-end OCR software packages, even when their costs exceed other packages, because their benefits in terms of fewer errors may outweigh the initial costs.

Vendor. Vendors charge around $1 - $2 per page to enter data via keyboard. Vendors charge similar rates for OCR text scanning, usually with a minimum charge of about $45 and a discount on a high volume of work. Vendor prices for OCR scanning will vary, depending on accuracy, volume, and the amount of markup in the original material. Image scanning costs vary from $0.60 - $3.50 per page, again, depending on volume and amount of data cleanup required.

Whether this step is performed in-house or by a vendor, it is recommended that microfiche not be planned for direct conversion because the technology still does not exist for doing this. Information on microfiche will require conversion to paper format before it can be converted, unless the original data are available for use.

3.2.3 Step 3 - Quality Assurance

In-house Staff. Checking the accuracy of scanned data can be performed in-house with basic word processing and computer skills. However, it is a very labor-intensive process; the greater the number of errors resulting from the data entry or scanning process, the more time will be required to clean up the files. Again, the consideration for performing this in-house is based on whether staff can dedicate sufficient time and the attention to detail required to attain the desired (or acceptable) level of data cleanliness.

Equipment. PC(s) with word processing software are needed to edit the data files for quality assurance.

Vendor. Quality assurance may be performed by a data conversion service vendor, although not all vendors do this. The cost of having an outside vendor perform this step may depend on both the length and content of the document. Documents containing multiple text fonts, foreign language text, special text characters or symbols, or captioned figures may cause unusually large numbers of conversion errors. In addition, documents containing special characters, abbreviations, handwritten annotations, or specialized scientific notation may be beyond the capability of an outside vendor to proofread for accuracy and may require a knowledgeable specialist for checking. Keep in mind that if a vendor is used, they have to be solely relied on to assure an error-free database. If a vendor performs quality assurance, it is generally included in the price for data conversion. When selecting a data conversion vendor, make sure to ask if quality assurance, and at what level, is included in the price.

3.2.4 Step 4 - User Interface and Indexing

In-house Staff. Because it is the developmental step that results in either a user friendly, appealing research medium or one that is cumbersome or dreary to use, the importance of

using experts for designing the user interface and indexing the material cannot be overemphasized. Whether the indexing program is designed in-house or a vendor's indexing program is used, the expertise of a designer who can set up search structures and organize data files is essential. It is recommended that a commercially produced indexing and retrieval software package be used. Many older commercial indexing programs are not user friendly and may require computer programming expertise. It is recommended that any indexing program be reviewed prior to purchase to determine its user friendliness for authoring and whether the software has clear instructions documenting how to use it.

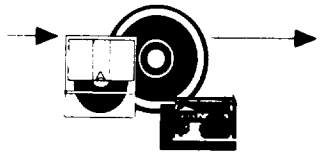
Indexing a database requires defining the data type, describing the data, and structuring the data for retrieval by providing appropriate specifications for the chosen output. Depending on the type of data (e.g., text, graphics), software should be chosen that supports the industry recommended standards for formatting data.

Indexing can be a time-consuming process. One recommended method of estimating the staff hours required for indexing is to divide the number of megabytes in the source file by four. Thus, a 400 megabyte database would require about 100 hours to index. Staff expertise and time demands, as well as the cost of the indexing software package need to be considered. Again, consultation with an experienced designer or vendor is recommended.

Equipment. An IBM-compatible PC (486, 33MHz or better) with 8Mb and a hard drive with a temporary storage capacity three to five times the size of the source database are needed to index the database. The amount of storage space needed will depend on the type of data format, (i.e., full text information or fielded databases). Full text information (books or other standard documents) requires less temporary disk space. Fielded databases (spreadsheets) require much more. As an example, if the database has 300 megabytes of text information, 900 megabytes of temporary disk space would be needed for the indexing process. If that 300 megabyte database were fielded information, perhaps 1500 megabytes of memory would be needed. If sufficient disk space is not available, it is possible to index in sections and save each section to magnetic tape. The finished, indexed product will require about 50% more disk space than the original source file; that is, a 400 megabyte database will increase in size to 600 megabytes. An indexing and retrieval software package is needed to structure the database. If in-house personnel do not have computer programming experience, commercial indexing and retrieval software should be purchased. There are many software packages commercially available that have a variety of capabilities and features. Once the package is purchased, indexing may be performed in-house. The choice of an indexing and retrieval software package is a critical step; the package should be user friendly and must fit all needs without having unnecessary and costly extras. Such software packages generally cost $1,000 - $4,000, with the higher cost reflecting the program's search capabilities and ability to automatically find and correct errors. Vendors also charge additional licensing and/or royalty fees for use of their software, which can run up to $30,000 for a flat fee in addition to a royalty fee paid per title, depending on duplication volume. (See Section 3.1.1, paragraph Licensing/Royalty Fees, for more detailed information.)

Vendor. Vendor prices for indexing can vary dramatically from $10,000 - $100,000 (not counting licensing and royalty fees), depending on the sophistication of the search system and the amount of data to be processed. (A vendor should be able to provide a price estimate based on these factors.) Just as selection of an indexing and retrieval software for in-house use is critical, so is ensuring that the vendor's indexing method meets NASA requirements in terms of search and retrieval capabilities, user friendliness, etc. Deciding whether to use a vendor to index data is made by considering how many indexed discs are needed now and whether others will be required in the future. If the user wishes to publish many CD-ROM's, it may be quicker and more cost-effective to purchase a software package and to index the material in-house with the assistance of a data processing consultant.

3.2.5 Step 5 - Transfer of Source Data to Magnetic Medium



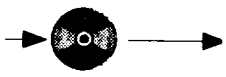
5. Transfer to Magnetic Medium

In-house Staff. If the vendor chosen for mastering does not accept the medium that now contains source data (e.g., computer floppy disks, hard disk), the data must be transferred to an acceptable magnetic medium. This can be performed in-house: basic computer skills and knowledge of the magnetic medium chosen are required. Again, the availability and experience of staff and the cost of necessary equipment are prime considerations in deciding whether to perform this step in-house.

Equipment. A 1-gigabyte (minimum) hard drive, magnetic medium hardware (e.g., 9-track tape or 8mm data cartridge drive) and tapes or cartridges are required. Nine-track tape drive prices are in the $5,000 range and 8mm data cartridge drives are about $2,000. Tape and cartridge costs start at about $20 each. If it is necessary to convert data, it is important to note that four to five 9-track tape reels are required to hold the data on one CD-ROM disc; 8mm data cartridges can hold the equivalent data of about three and a half CD-ROM's.

Vendor. There are many service vendors who will transfer data from computer floppy disks to a magnetic medium. On average, for transfer from floppy disk to magnetic tape, vendors charge about $10 per floppy disk (plus the cost of the magnctic tape if the customer does not supply it). Again, it is important to remember that one CD-ROM disc can hold up to 450 high-density floppy disks. Therefore, using a vendor for this step can become quite expensive if a large volume of data needs to be transferred. However, almost all mastering facilities now accept cartridges and removable drives as data input media, so it should not be necessary to transfer source data stored on floppy disks.

3.2.6 Step 6 - Premastering



6. Premaster CD-ROM

Once the data, documentation, and software files have been prepared and validated, the procedure of creating the CD-ROM's can begin. This process involves premastering, creating and testing proof discs, and eventually generating the final CD-ROM's. It is a somewhat complicated and potentially time-consuming process which offers some hardware/software configuration options which can that affect the cost of the work.

The first step of the premastering/mastering procedure, which is so highly recommended as to be a requirement, is to create some proof (or one-off) discs for testing purposes. A proof disc is a temporary copy of the data that is cheaper than the entire mastering process of creating the glass master and generating replicates. There is no artwork on the proof disc, but it contains, in CD-ROM format, all of the data that was written to the work hard drive.

To create proof discs, it is necessary to premaster the selected data. Again, premastering is the process of converting the prepared data into a form that can be written onto discs. Premastering is performed by software that can write the converted data to hard disk, magnetic tape or directly to the proof disc, depending upon the hardware/software configuration available.

After the proof disc is reviewed and required changes are made, either a second proof disc can be generated from the revised data or the final mastering can be performed. If the second proof disc is generated, a review would again take place, resulting in another version. More proofs and versions could be generated until confidence is reached that the discs are "complete, accurate, and usable." When the decision is made to do the final mastering, the final version of the data (along with booklets, inlay trays, and disc art) are sent to the mastering facility, which creates the master disc. There, all of the replicates are packaged in plastic "jewel boxes" and covered in shrink-wrap plastic. There are three basic ways in which this procedure can be accomplished, each with its own costs, advantages, and disadvantages.

In-house Staff. Premastering involves converting the database into CD-readable files (formatting block address, headers, and sync pattern), logically placing the files on the disc (ISO 9660), and inserting error detection and correction codes into the database. The CD-ROM reader then uses these codes for accurate retrieval. Premastering is a relatively fast process, requiring two to five hours.

Equipment. IBM compatible PC (486, 33MHz or better), either, 1) desktop encoder/recorder system—sometimes called a write-once system—(about $30,000) and write-once discs ($30-$80 each), or 2) a CD-R system ($2,000 - $5,000), CD-R disks (about $15 each), and a premastering software program.

Vendor. Many vendors premaster data at flat prices ranging from $250 - $750. Others charge about $125 per hour, with an average time required of two to five hours. Prices will vary depending upon the input medium (e.g., computer hard drive, magnetic tape, or cartridge). Only vendors who will format the database to the ISO 9660 standard should be selected.

3.2.6.1 Premastering Onsite

The procedure to purchase software to perform the premastering step and use an outside mastering service to produce the CD-ROM's follows these steps:

1. Purchase the premastering software and arrange for the services of a mastering service (for one-offs and final replication).

2. The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) to a magnetic tape. It is preferable to be able to write all of the data on the directory structure to one tape. Determine what media is acceptable by the premastering service.

3. The mastering service will take the data tape and create a one-off disc which will be sent back for disc review.

4. After changes are made to the original disk files in response to the review, a new magnetic tape is written and sent to the mastering service. Disc artwork and a booklet and inlay tray, if created, are also sent to the mastering service.

5. The mastering service creates the master disc and as many replicates as were ordered and sends them back.

Premastering onsite is usually advisable if more than a couple of discs will be produced.

3.2.6.2 Premastering One-offs Onsite

The most sophisticated procedure available is to purchase hardware and software to perform the premastering step and to create draft CD-ROM's (one-offs) in- house and then use an outside mastering service to produce the final CD-ROM masters and replicates. Because of the decreasing prices of the machines ($2,000-$5,000) that produce the one-off discs (which cost approximately $15 each) and the rapid review copy turnaround time, this has become the most cost-effective and efficient means of premastering CD-ROM's. The following steps outline the in-house production of one-off CD-ROM's:

1. Purchase the premastering hardware (CD-R) and software and arrange for the services of a mastering service (for mastering and replication).

2. The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) directly to a one-off disc. As many discs as are necessary for review can be generated in-house.

3. After changes are made to the original disk files in response to the review, a new one-off disc is written and sent to the premastering service. Disc artwork and a booklet and inlay tray, if created, are also sent to the premastering service.

4. The mastering service creates the master disc and as many replicates as were ordered and sends them back.

Premastering one-offs onsite is usually advisable if several discs are being produced and more are anticipated.

3.2.6.3 Premastering Offsite

The procedure to use an outside premastering service to produce the CD-ROM's requires the purchase of no in-house software or hardware for CD-ROM production purposes and follows these steps:

1. Arrangements are made to purchase the services of a premastering service.

2. The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) to a magnetic tape using a utility that should be compatible with utilities available at the premastering service. It is preferable to be able to write all of the data on the directory structure to one tape. Determine what media is acceptable by the premastering service.

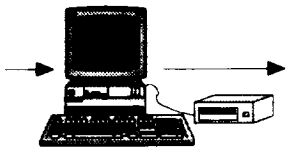
3. The premastering service will take the data tape and create a one-off disc which will be sent back for disc review.

4. After changes are made to the original disk files in response to the review, a new magnetic tape is written and sent to the premastering service. Disc artwork and a booklet and inlay tray, if created, are also sent to the premastering service.

5. The premastering service premasters the data to a one-off disc and sends the disc and the artwork to the CD-ROM mastering facility that creates the master disc and replicates.

All of the costs are paid to the premastering service, which pays the mastering facility for its services. Premastering offsite is usually advisable if only a couple of discs are involved and no more are anticipated in the foreseeable future.

3.2.7 Step 7 - Checking the Search Structure for Proper Retrieval (CD-ROM Simulation)



7. Check the Search Structure (CD-ROM) Simulation

In-house Staff. Following off-site premastering, the vendor will send a check disc to the client for quality assurance. The check disc simulates the CD-ROM disc so the client can access the software and perform searches to make one final check of the product's accuracy before the final master disc is pressed. For onsite one-offs (CD-R), the same checks would be performed as for off-site.

Note: This quality assurance step requires time and expertise in search structure and retrieval methodology.

Equipment. PC and a CD-ROM reader are needed to run the check disc.

Vendor. It is not recommended that a vendor perform this step because it is up to the client to determine if the program performs satisfactorily.

3.2.8 Step 8 - Mastering and Duplication

3.2.8.1 Master Disc Creation

8. Master and Duplicate CD-ROM Discs

Mastering must be performed by a vendor. It requires very specialized, costly equipment and a specialized environment. However, as CD-ROM technology gains wider industry acceptance, it is conceivable that within five years equipment capable of producing CD-ROM master discs will become commercially available and feasible for individual users to acquire. Turnaround time for mastering can range from 15 days to same-day service, with prices increasing for faster delivery time. Prices are generally about $800 for 15-day turnaround to $2,900 for 1-day turnaround. The price may also vary based on the input medium.

3.2.8.2 Vendor Selection

When selecting a vendor for the mastering process, consider the following: What types of input media does the vendor accept? Does the company have an integrated manufacturing line (pressing, checking for errors, labeling and packaging performed in one continuous line)? Ask about error and rejection rates and see figures if possible.

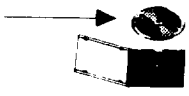
3.2.8.3 Master Disc Storage

Many vendors include one year's storage of the CD-ROM master disc and input media as part of their service. The storage service generally includes one free remastering if failure of the master disc occurs during storage. Additional storage time costs about $200 per year. A reorder charge may range from about $200 for a 15-day turnaround to about $300 for 3-day turnaround. About 50 replicas are generally included in this price. Some vendors waive the reorder fee and charge only for the discs duplicated.

3.2.8.4 Duplication

Duplication of discs after mastering involves a similar time range for delivery; prices are sometimes tied to the mastering charges. Generally, small quantity (100 - 200) replicas cost about $2 per disc. The price decreases with volume and increases for fewer copies or faster delivery time.

3.2.8.5 Labeling and Packaging



Package Final Product

Most vendors include one or two color labeling and bulk packaging in their mastering and replication prices. The labels are printed from positive, color-separated film. Additional colors in the labels can cost from $0 - $100 per color. For disc packaging, most vendors use the clear-plastic "jewel box," costing about $.35 per box. For $.10 to $.15, each disc can be packaged in an envelope, a clear plastic bag, or a cardboard sleeve.

3.2.9 CD-Recordable: AKA Home Cooking

CD-R provides an excellent tool for producing limited run CD-ROM's and premastered one-offs created onsite. The process of producing a CD-R is a subset of mass production of CD-ROM's with limits to the final number of CD's produced. Many vendors readily provide the necessary tools (hardware and software) to gather, arrange, index, test, produce, and design packaging inserts and/or jewel case covers. The steps required to this point are the same for both processes. CD-R offers the benefits of rapid turnaround and customization with limited production cost.

3.2.10 Other Considerations for In-house Production

3.2.10.1 Staff

Probably the most important consideration for in-house production of CD-ROM's is the availability of adequate staff. This means not only numbers of persons, but their ability to dedicate sufficient time to the project(s), their having or obtaining the expertise required, and the relative permanence of staff members with that expertise.

3.2.10.2 Staff Training

Vendors offer various software programs to aid in CD-ROM production; free training is provided with a small number of these programs. Some vendors also offer training workshops and seminars on CD-ROM production; and there are both commercial and academic programs in some disciplines. Vendors must be contacted to find out more about these services.

3.2.10.3 Quantity of CD-ROM's to be Produced

A few years ago, in-house production of one or two CD-ROM's would not be economically feasible unless all required resources—staff expertise, time, equipment—were already in place. However, the introduction of reasonable cost CD-R equipment has made it possible to create one-offs and very small run, specialized CD-ROM's. If a large number of CD-ROM's will be produced and production will continue indefinitely, initial investments in equipment, software, and training will prove to be less costly than repeated purchases from vendors.

3.2.10.4 CD-ROM Market

Whether CD-ROM production is performed in-house or by a vendor, consider that there may be a market for any CD-ROM's produced by the NASA STI Office.

3.2.10.5 Emerging Standards

There are many emerging standards that govern or may govern the production and publishing of CD-ROM. As potential producers entering the market, NASA STI Office

personnel must be knowledgeable of these standards and be able to adopt them as they become available or as they are updated.

3.2.10.6 General Implementation and Maintenance

The timeframe and cost for implementing a CD-ROM system will vary depending on the state of the source data (paper, computer disk, microfiche, film, video, magnetic tape), the quantity of data, and the vendors used for various stages of the CD-ROM production. However, as the technology for CD-ROM improves and evolves, the associated costs and timeframe for production and implementation will change to reflect the advances in hardware, software, and vendor services.

With any technology or hardware, there are maintenance requirements that must be considered. Concerns include design factors to be considered before CD-ROM system installation and maintenance strategies to be followed after installation.

Design decisions made before system installation can ensure that the system is maintainable for years after it is installed. Basic guidelines for a multi-user CD-ROM information retrieval system include:

Modularity and Interchangeability. A system built from many interchangeable modules is inherently more reliable and maintainable than a system based on a single central device, as long as the rest of the system can continue to function if one module fails. An information retrieval facility based on a network, with PCs or other compatible computers exchanging data on the network, meets this goal. PCs can be added, repaired, or replaced as necessary without disabling the system as a whole.

Vendor Independence. A key to long-term maintainability is freedom from dependence on a single vendor. Use of interchangeable parts is of primary importance. Rather than committing to a single vendor of computer hardware, CD-ROM drives, or development software, a sounder strategy is to make sure that any hardware or software used in the system is compatible with multi-vendor standards for operation. This will help ensure that even if the vendor of a component used in the system goes out of business or discontinues the product, a replacement from another vendor will be available.

Standards. Another key to long-term system maintainability is the use of standards that are now being internationally recognized.

Bibliography

— Angelici, G. and Skiles, J.W. 1993. *Publication of Science Data on CD-ROM: A Guide and Example*. NASA Ames Research Center. NASA Contractor Report 4560.

— Fricks, J.R. 1992. *Compact Disc Terminology*. Disc Manufacturing Inc.

— King, J.H. and Grayzeck, E.J. 1989. *Minutes of the CD-ROM Workshop*. NASA Goddard Space Flight Center. NASA Technical Memorandum 105066.

— Redmond, K., Sheppard, J., Humphrey, M., Lee, S. 1992. *CD-ROM Applications in Professional Military Education*. Naval Personnel Research and Development Center. NPRDC-TN-93-1.

Attachments

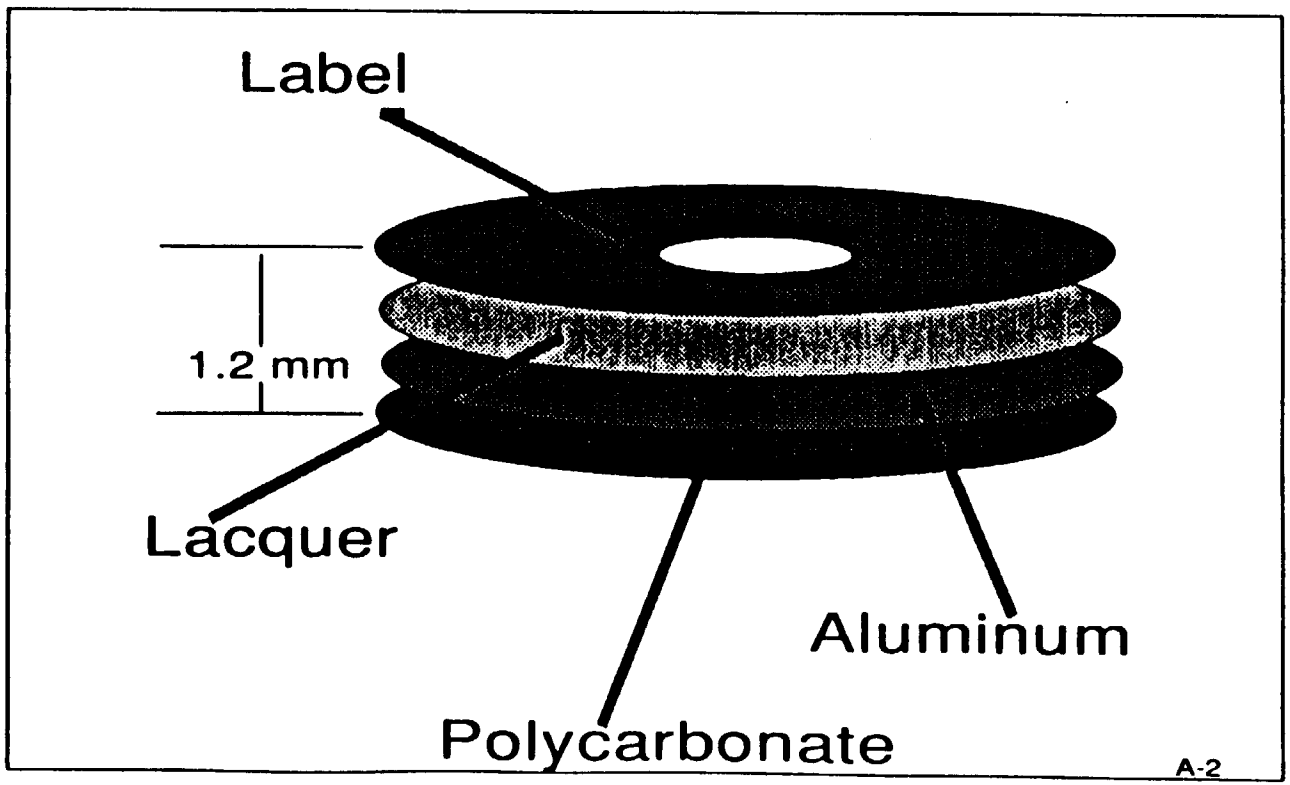
(Attachments 1-4, 11 courtesy of the Institute for Federal Printing and Publishing, United States Government Printing Office)

Attachment 1

CD-ROM Physical Characteristics

PHYSICAL CHARACTERISTICS AND FILE FORMAT

- Disc measures about 4 3/4 inches diameter
- About 1.2 millimeters thick
- First layer is the label
- Second is the protective coating of lacquer
- Third layer is reflective coating of aluminum
- Fourth is the polycarbonate plastic substrate containing pits and lands



CHARACTERISTICS, STANDARDS, AND RETRIEVAL CONCEPTS

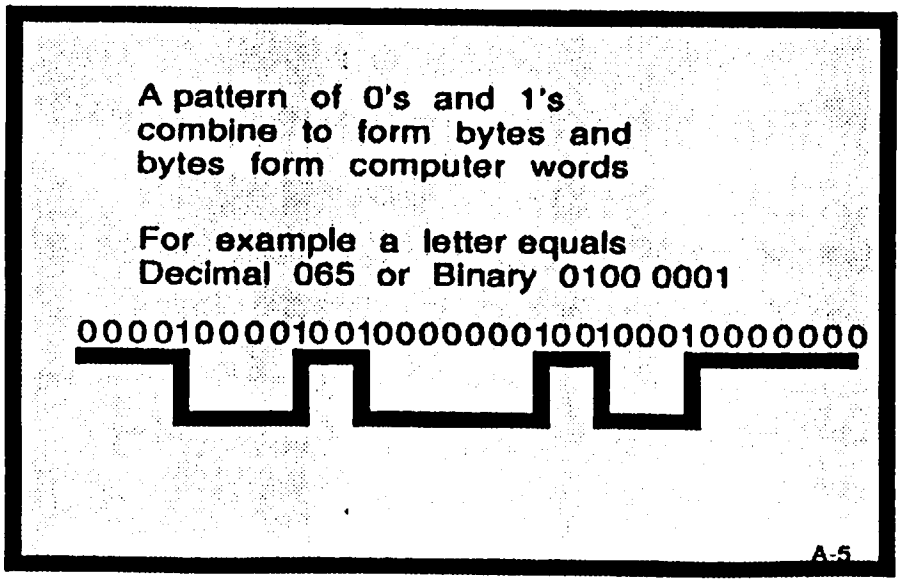
- Physical and logical format describe the pattern in which data is represented on the disc
- Both formats are standardized and guarantee that disc will be readable
- The physical format was devised by Philips and Sony (ISO 10149)
- The logical format was created by a committee of all manufacturers (ISO 9660)

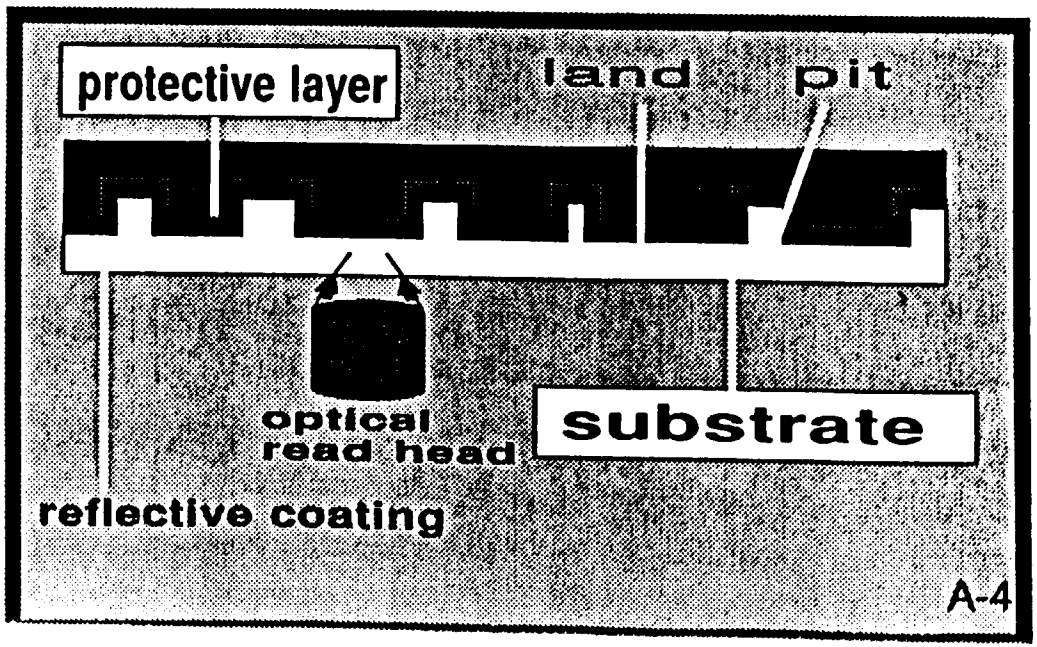
Attachment 2

CD-ROM Physical Characteristics

PHYSICAL CHARACTERISTICS AND FILE FORMAT--Continued

- Track containing the pattern consists of a single spiral
- Starting from the center, up to 20,000 tracks from center to edge
- The Track contains tiny "pits" and flat surfaces called "lands"
- 99 tracks will fit on a CD-ROM
- The pits and lands cause a variation in light that registers ones and zeroes
  When the surface changes from flat to pit a "one" is recorded
  Zeroes are constantly being interrupted by "ones" to create
  the binary word values
- Constant Linear Velocity must be maintained for readability





ACRETCD2

PHYSICAL CHARACTERISTICS AND FILE FORMAT--Continued

- The sectors on the disc are structured with 12 bytes of synchronization code; 4 bytes for address or name; 2048 bytes of data; and 288 bytes of error detection and correction for monitoring the reading of data
- Error detection and correction codes are much more efficient than audio discs
- A mathematical checking system or algorythm is used to make sure no incorrect data can be read from the disc
- Data is blocked and sums are recorded to double check and verify what is being read
- This checking process can actually correct misread bytes of data
- The system will acknowledge a failure to read accurately

Attachment 3

How To Embark On A CD-ROM Endeavor

HOW TO EMBARK ON A CD-ROM ENDEAVOR
USING SEARCH SOFTWARE

WHAT YOU SHOULD KNOW ABOUT SEARCH SOFTWARE BEFORE EMBARKING ON A CD-ROM ENDEAVOR

AA2DYROM

WHY DO I NEED TO KNOW?

- SAVE MONEY
- SAVE TIME
- MAKE INFORMED DECISIONS
- PRODUCE A BETTER PRODUCT

- 40% TO 70% OF THE TOTAL COSTS ARE IN DEVELOPMENT

WHAT HAPPENS FIRST

- FUNCTIONAL ANALYSIS
- DATA ANALYSIS
- CONCLUSIONS
- SOFTWARE SELECTION
- DATA PREPARATION
- SOFTWARE DEVELOPMENT
- DISC PROTOTYPE
- DISC PRODUCTION

FUNCTIONAL ANALYSIS

- WHO WILL USE THIS PRODUCT
- HOW WILL THEY USE THIS PRODUCT
- ARE THERE ANY SPECIFIC FEATURES REQUIRED
    - * Special Search Screens
    - * Special Screens
    - * Images

DATA ANALYSIS

- WHAT TYPE OF DATA
- HOW MUCH ALTERING WILL BE REQUIRED
- QUANTITY OF DATA

CONCLUSIONS TO THE FUNCTIONAL AND DATA ANALYSIS

- TYPE OF SOFTWARE TO BE USED
- SOFTWARE CONFIGURATIONS
- TIME AND COST ESTIMATES

TYPES OF DEVELOPMENT SOFTWARE

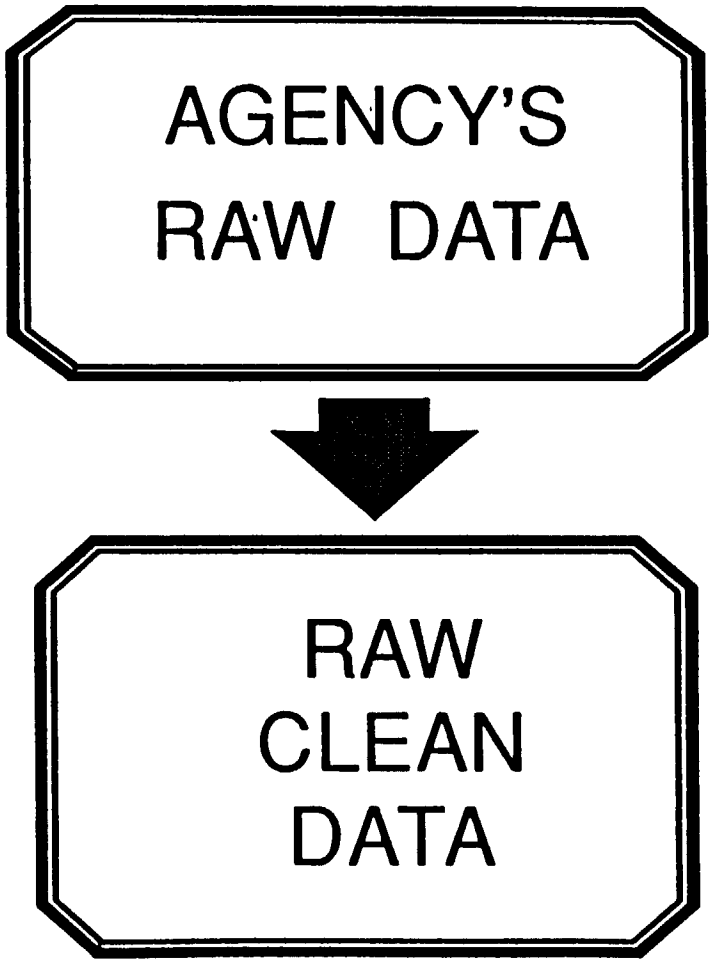
- RECORD PRODUCTS
- TEXT PRODUCTS

NOTE: SOME SOFTWARE VENDORS COMBINE THESE DATABASE TYPES AS ONE SOFTWARE PACKAGE

DATA PREPARATION

- RECORD
  - TYPE OF RECORD FORMAT
    - COMMA DELIMITED
    - FLAT DATABASE
  - FIELD TYPES
    - NUMERICAL
    - DATE
    - STRING
    - TEXT
- TEXT
  - STRUCTURE
  - SPECIAL FIELDS

QUICK OVERVIEW OF CDROM DEVELOPMENT

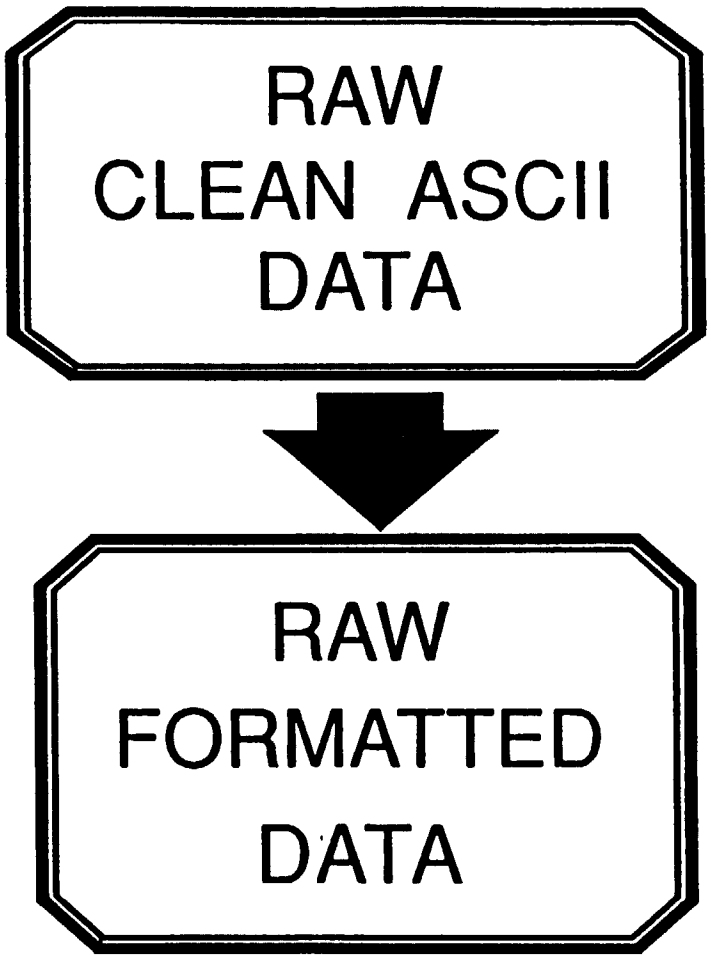


GENERAL CONVERSIONS:

EBCDIC TO ASCII CONVERSIONS

WORD PROCESSOR TO ASCII

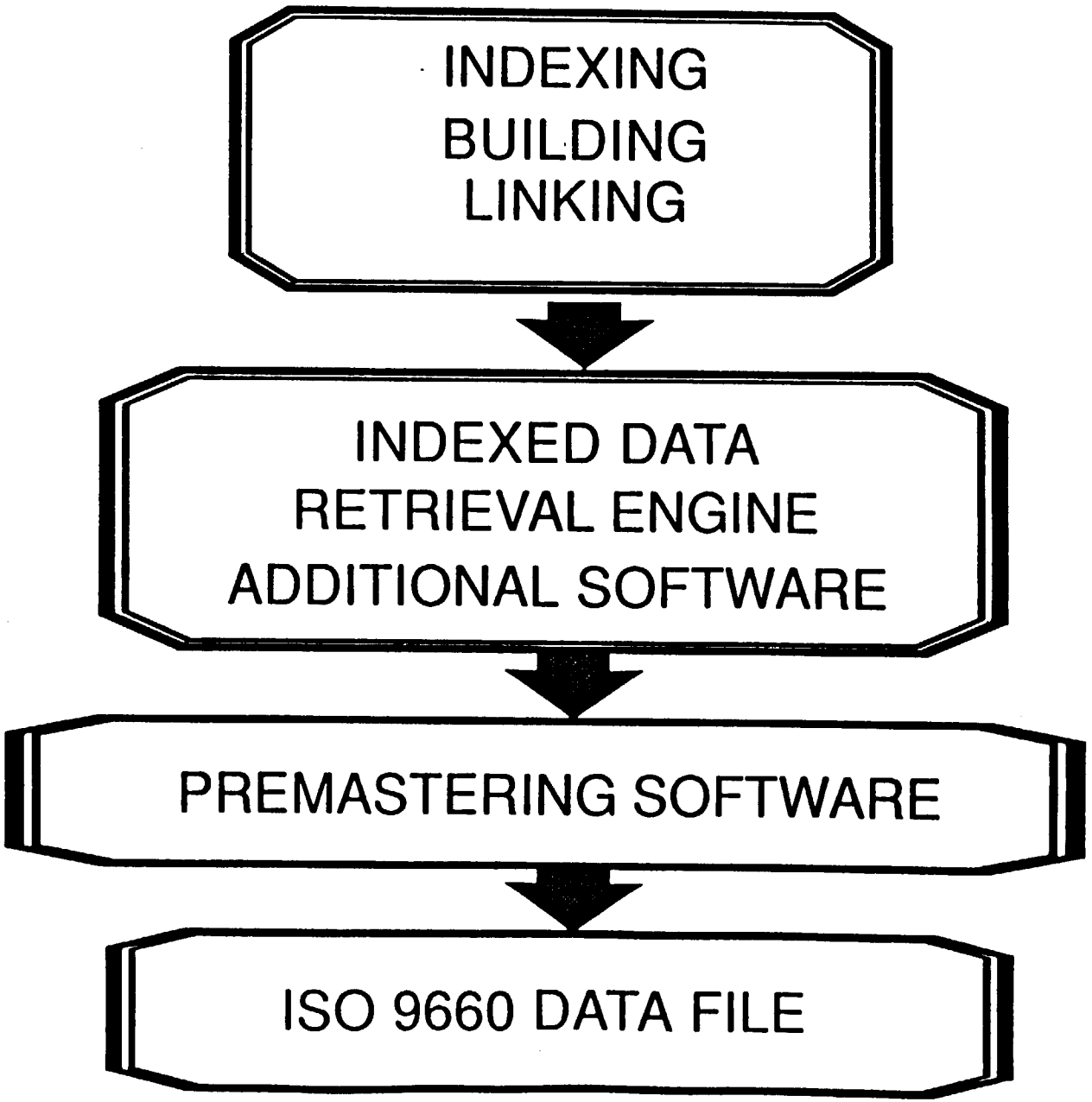
QUICK OVERVIEW OF CDROM DEVELOPMENT--Continued



THIS STEP USUALLY CONSUMES THE MOST TIME AND MONEY

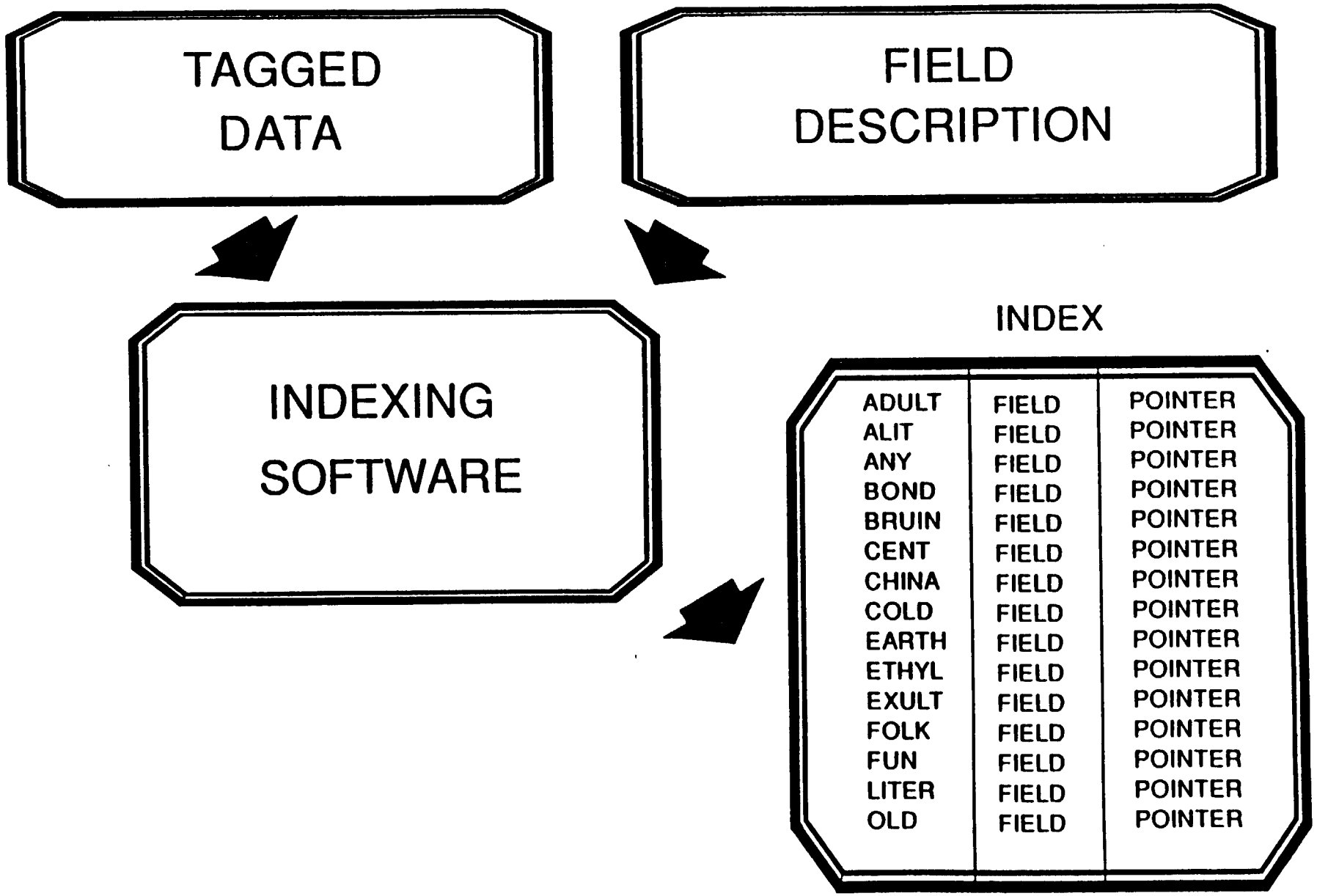
AJI2DYROM

QUICK OVERVIEW OF CDROM DEVELOPMENT--Continued



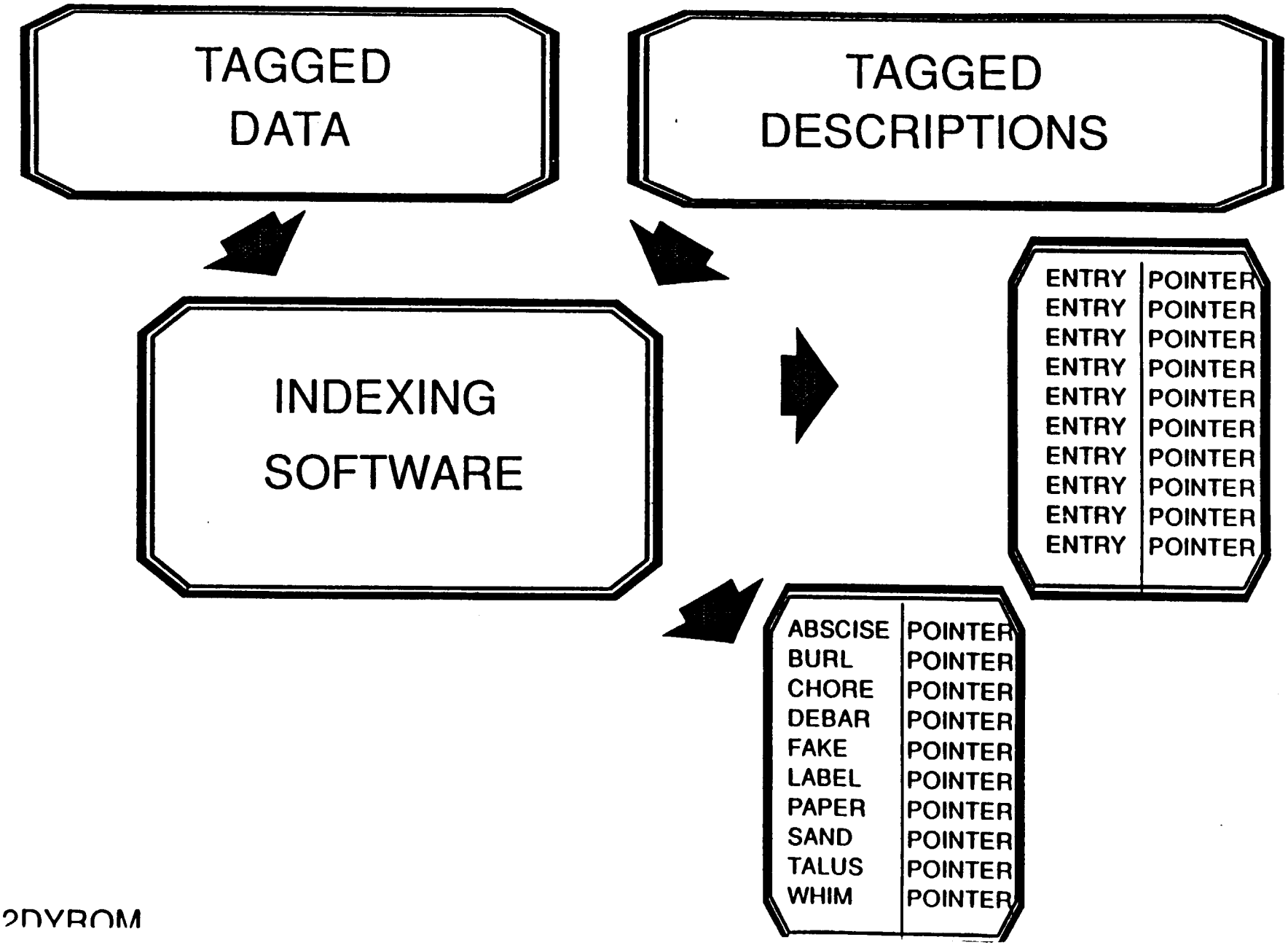
AK2DYROM

RECORD INDEXING



AL2DYROM

TEXT INDEXING



AM2DYROM

MATERIALS THAT CAN BE PROVIDED TO SAVE TIME AND MONEY?

- LABEL ALL MATERIALS (TAPES, DISKS, AND COPY)
- SUFFICIENT DOCUMENTATION
- AVOID SENDING DATA PIECEMEAL
- PROVIDE CONTACT NAMES, NUMBER AND FUNCTION
- HARD COPY OR PUBLICATIONS

RECORD DATABASE

- PROVIDE A CLEAN DATA BASE
- BYTE COUNT DELIMITED FIELDS
- FIELD LENGTH DELIMITED FIELDS
- PROVIDE THE DATA IN ASCII FORMAT
- COMMA DELIMITED FIELDS
- A LIST AND OR LOCATION OF UNUSUAL ASCII CHARACTERS

RECORD DATABASE

- PROVIDE A DETAILED "MAP" OF THE DATA BASE RECORDS
- RECORD LENGTH OR TERMINATING CHARACTER
- FIELD LABELS (ONE OR TWO WORD LABELS)
- CHARACTER COUNT OR MAXIMUM CHARACTER COUNT FOR EACH FIELD
- FIELD TYPE LABELS; STRING, TEXT, NUMERIC, DATE, ETC.
- FIXED LENGTH FIELDS PADDED OUT WITH SPACES (ASCII 32)

TEXT DATA BASE

- PROVIDE A CLEAN DATA BASE
- PROVIDE A DETAILED "MAP" OF THE DATA BASE
- STRICT RIGID STRUCTURE
- TAGGED DATA STRUCTURES
- TABLE OF CONTENTS AND INDEXES
- A LIST AND OR LOCATION OF UNUSUAL ASCII CHARACTERS

GRAPHICS

- PHOTOS
- COLOR IMAGES
- FIGURE REFERENCES
- TABLES
- SPECIAL CHARACTER TEXT

Pre-Scanned Images Can Be Provided

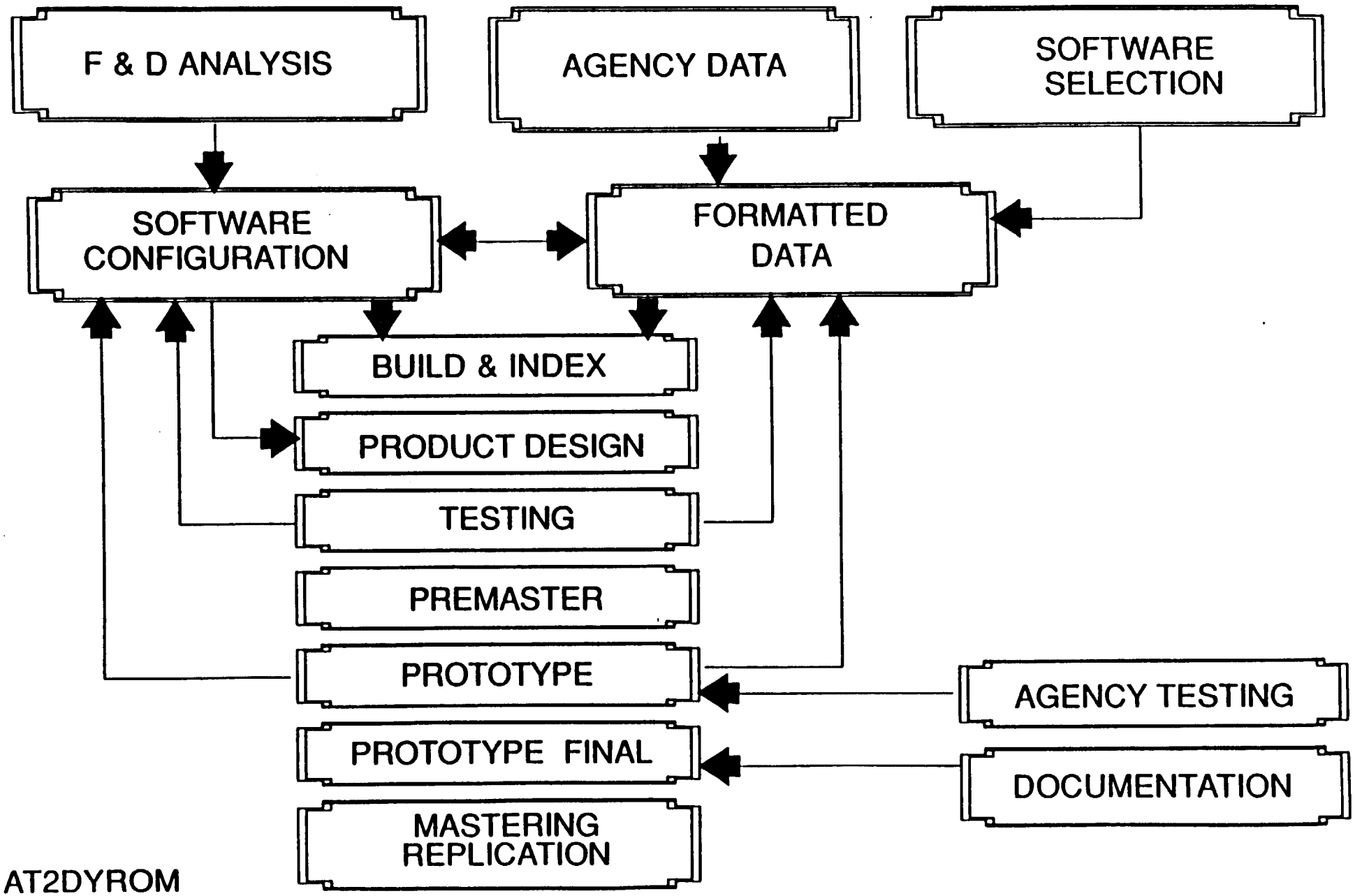
A Picture is Worth a Thousand Words

GRAPHICS

- SCANNED IMAGE FORMAT
  - TIFF FORMAT
  - PCX FORMAT
- CLEAN COPY (IF SCANNING)
  - PHOTOS
  - CAMERA COPY
  - ORIGINAL DRAWINGS
  - SINGLED SIDED
  - DARK COPY
- MAP OF IMAGE LOCATION IN TEXT (TAG AND LABEL)
- WELL LABELED IMAGES
- 7 TO 15 MINUTES PER IMAGE IS AVERAGE TIME REQURED
  HALF TONE SCANNING TAKES LONGER
  FRONT AND BACK IS TWO SEPARATE IMAGES

AS2DYROM

FLOW



AT2DYROM

Attachment 4

Steps Required To Develop A CD-ROM

STEPS REQUIRED TO DEVELOP A CD-ROM

FIVE MAJOR PHASES THAT OCCUR

- Data Development and Conversion
- Structuring the Text and Coding
- Building the Indexes for Searching
- Writing the Data to ISO 9660 Format for Replication
- Premastering and Mastering Processes

DATA DEVELOPMENT AND CONVERSION

- CD-ROM data must be machine readable

- Data may exist as:
  - typewritten manuscript
  - electronic data containing control codes
  - files coded with composition language
  - images of text pages

- Manuscript, microfilm and other non-electronic data must be re-keyed or scanned

- Scanning forms, text, or drawings will create a single unit or element
  OCR scanning will result in searchable character data

DATA DEVELOPMENT AND CONVERSION--Continued

- Re-keying the data is often the most sensible approach to developing a database

- Here are some ways to avoid the hidden costs that begin to accumulation after a project begins:

  be sure to understand how the charges are calculated for the keystroke count (hundreds? thousands?)

  know what medium will be used to deliver the database, tape, disk, etc.

  test samples before making proceding with project

  consider the need to add tags to data as it is keyed

  avoid being ask to pay more as problems are encountered

DATA DEVELOPMENT AND CONVERSION--Continued

check the size of halftones and borders to be sure they will fit on laser printers

verify keystroke accuracy level required

give consideration to having a project manager

be sure to use a reliable service bureau with a good reputation

- A magnetic tape or disk containing character-coded data is the best form especially when free from unwanted processor codes
- Software "bridge" programs often can successfully filter data into the form desirable for ISO 9660 format
- 40% percent or more of the total costs to publish may be spent in developing and structuring the data and building the indexes

TEXT PREPARATION

- Data must be free of nonessential hard codes
- Tags or markers have to be inserted to indicate:
  - New chapters
  - Subtitles
  - Elements to be searched
  - Footnotes
  - Identitfiers required for Hypertext searches
  - Codes that assist in screen display
- Before choosing a software retrieval system:
  - Evaluate the structure of the data
  - Consider the appearance on the screen when retrieved
  - Understand how data elements need be identified for retrieval

TEXT PREPARATION--Continued

- To display data on a viewing monitor or print the data there are generally three types of retrieval systems:
  - One system requires inserting special characters in the data for centering or indenting, and other formatting needs
  - Another system accomplishes the task by inserting general instructions and permitting the display program to determine how the elements will appear
  - And a third system uses a style table or lookup list to reference how each element should appear

- All the steps to develop data must be accomplished with an eye focused on the requirements of the software retrieval system selected when the projected was first begun

STANDARD GENERALIZED MARKUP LANGUAGE

Not: STANDARD GRAPHICS MARKUP LANGUAGE

- Standard Generalized Markup Language (SGML)
- International Standards Organization (ISO) Standard for Document Description (ISO 8879)
- A solution to to document description and structure language that is made for humans and computers to use without difficulty
- The Department of Defense through the CALS initiative has been a leader in implementing SGML
- The American Association of Publishers, the airline industry, General Motors, and others requiring reliable documents
- SGML is a system used to define the elements in a document, their relationship to the document itself, and content meaning

SGML--Continued

- Definition of elements in SGML is independent and may be applied to any operating system, format, or computer
- Text preparation often demands element identification for search and retrieval and SGML is more than capable of achieving this coding structure
- Most data at best will contain procedural markup used for typesetting which only describes the appearance and position, of data elements and how they appear on the printed page
- Descriptive element tags work quite differently These tags identify data elements in terms of their meaning, content, relationship to other elements in the document
- Merging databases with similar tagging schemes or simply tagging data for retrieval, usually requires a markup phase that can be costly and time-consuming

SGML--Continued

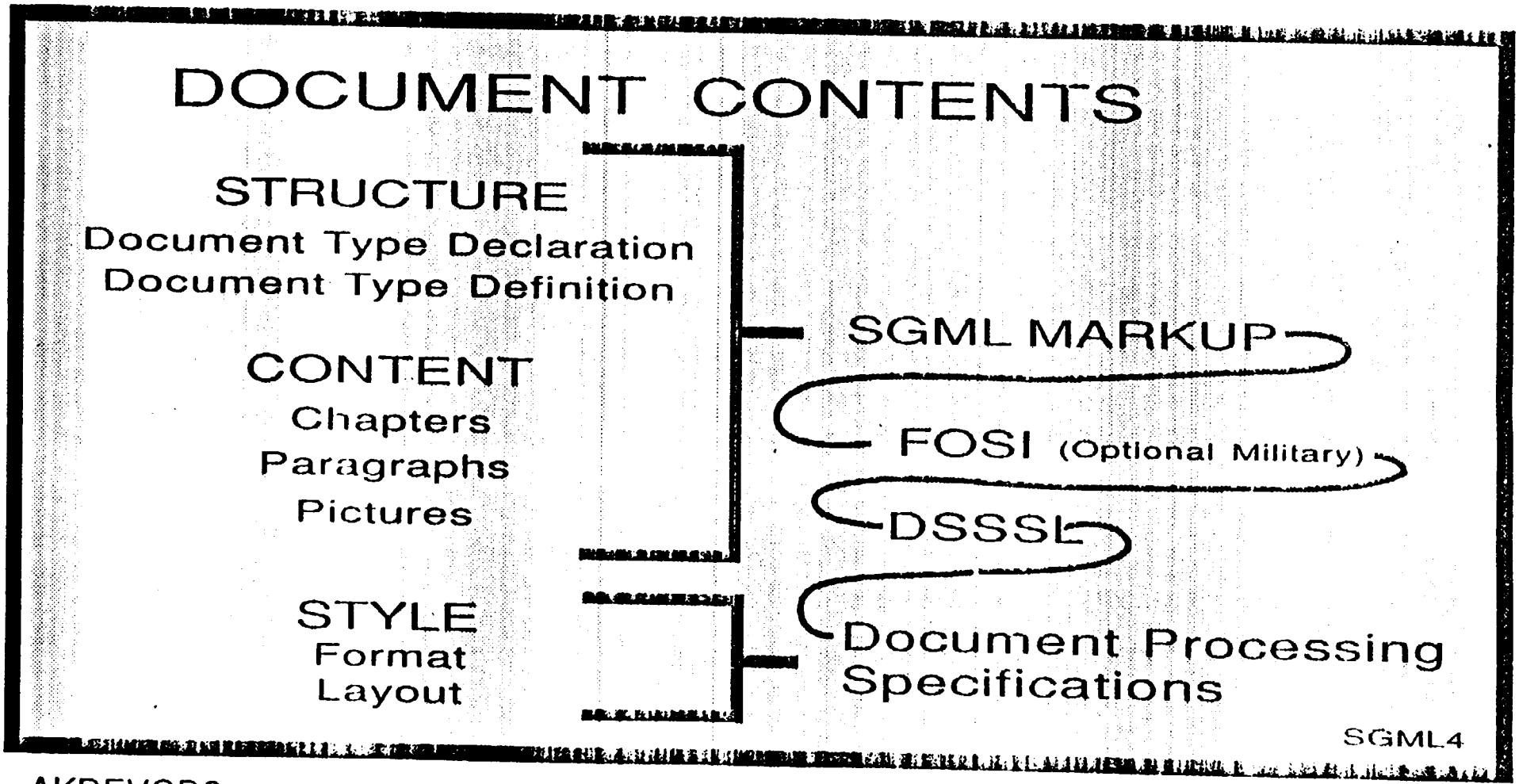
- Full-text tagging offers many advantageous worth consideration. Tagging for cross-referencing, indexing, hypertext searching, defining a zone as a search block
- Tags can be used for special functions, too. Hidden tags can permit searching by topic fields that are not obvious on the viewing screen. "No search" tags can render large segments of data unsearchable thus saving time. Title tags and data fields are also possible tagging elements
- SGML will also describe the hierarchical structure of chapters, subchapters, quoted material, and footnotes
- Tags that affect how the data appears on the viewers screen in bold, underline, inverse mode or other modes such as blinking can be related to SGML tagging schemes

SGML--Continued

- SGML is a markup scheme for defining document data elements
- SGML is not a standard set of tags for a specific procedure
- SGML is a way to describe or create documents independent of hardware, software, formatter, etc.
- SGML is a set of identifier tags for describing elements and their relationship to the document
- SGML identifies two of the three typical characteristics found in a document: (1) structure; (2) content; but not (3) style

SGML--Continued

- The SGML-structured document can be filtered through special programs for applications requiring special coding



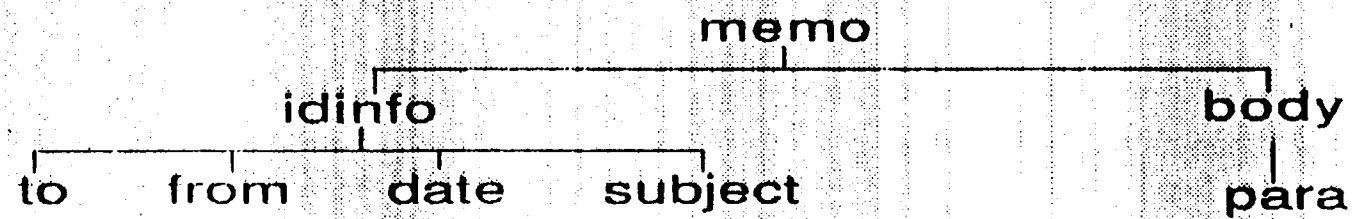
AKDEVCD2

SGML--Continued

- SGML5 A structure chart can be useful to help become familiar with SGML and also to analyze a document for content

STRUCTURE

Architecture of document, not format or layout



Document Type Definition (DTD)

```
<!DOCTYPE memo  [
<!ELEMENT memo - - (idinfo, body)>
<!ELEMENT idinfo - - (to, from, date, subject)>
<!ELEMENT body - - (para)>
<!ELEMENT to - - (#PCDATA)>
<!ELEMENT from - - (#PCDATA)>
<!ELEMENT date - - (#PCDATA)>
<!ELEMENT subject - - (#PCDATA)>
<!ELEMENT para - - (#PCDATA)>
] >
```

SGML5

SGML--Continued

- SGML6 The word "instance" in SGML vocabulary refers to the document with the tags imbedded in it
- The instance can be "parsed" or checked for valid composition by checking it against the DTD. This process of validating is called "parsing"

SGML Instance

```
<memo>

<idinfo>

<to>John Doe</to>
<from>Mary Smith</from>
<date>June 30, 1978</date>
<subject>Personal Holiday</subject></idinfo>

<body>

<para>All personnel are entitled to one personal day
per calendar year.</para>

</body></memo>
```

SGML6

INDEX BUILDING

- The database on the CD-ROM is usually too large to be searched quickly and efficiently without a search scheme.
- Special software containing a "build engine" and a "search engine" must be applied to the coded database
- The "build engine" organizes the data by indexing the searchable words prior to premastering
- The indexes are called inverted because they are alphabetically arranged word lists with "pointers" that map the location of each word in the data
- The indexes, data, and auxiliary files are all that is necessary for the "search engine". The "build engine" is no longer needed and is not written to the CD-ROM

INDEXING BUILDING--Continued

- The "search engine" is the software that interprets requests and seeks information requested
- Full-text data and fielded data indexes are created with a separate "build engine" and must be discussed separately
- Full-text indexing procedures list every searchable unit (words and phrases) in the database. Exceptions to this are words that are of no consequence and appear too often
- Full-text indexes contain every searchable work in alpabetical order
- The indexing system can be several layers deep with the top layer serving as the dictionary
- The dictionary points to a reference index file where each occurrence of the word being searched listed for selection

INDEXING BUILDING--Continued

- Fielded data must be indexed separately and with a "build engine" designed specifically for fielded data files
- Fielded data is found in records of a specific length. Each record contains these fields in the same order from record to record
- Software for this purpose varies from company to company in how it recognizes fielded elements. One program requires by commas, while others require specific length for the fields
- Fielded documents have more powerful search capability and are faster in responding to search requests (this is called seek time)
- Data containing items such as: name, address, zip code, etc. are usually considered for fielded structure

INDEX BUILDING--Continued

- Full-text or fielded build engines both create the dictionaries and reference files required to perform retrieval
- The input files needed for writing to tape and forwarding to the mastering facility are composed of the searchable database in correct format, compressed scanned images, indexes, retrieval display software, and auxiliary files. These files must be placed in the order in which they are to reside on the disc

WRITING DISC IMAGE TO TAPE

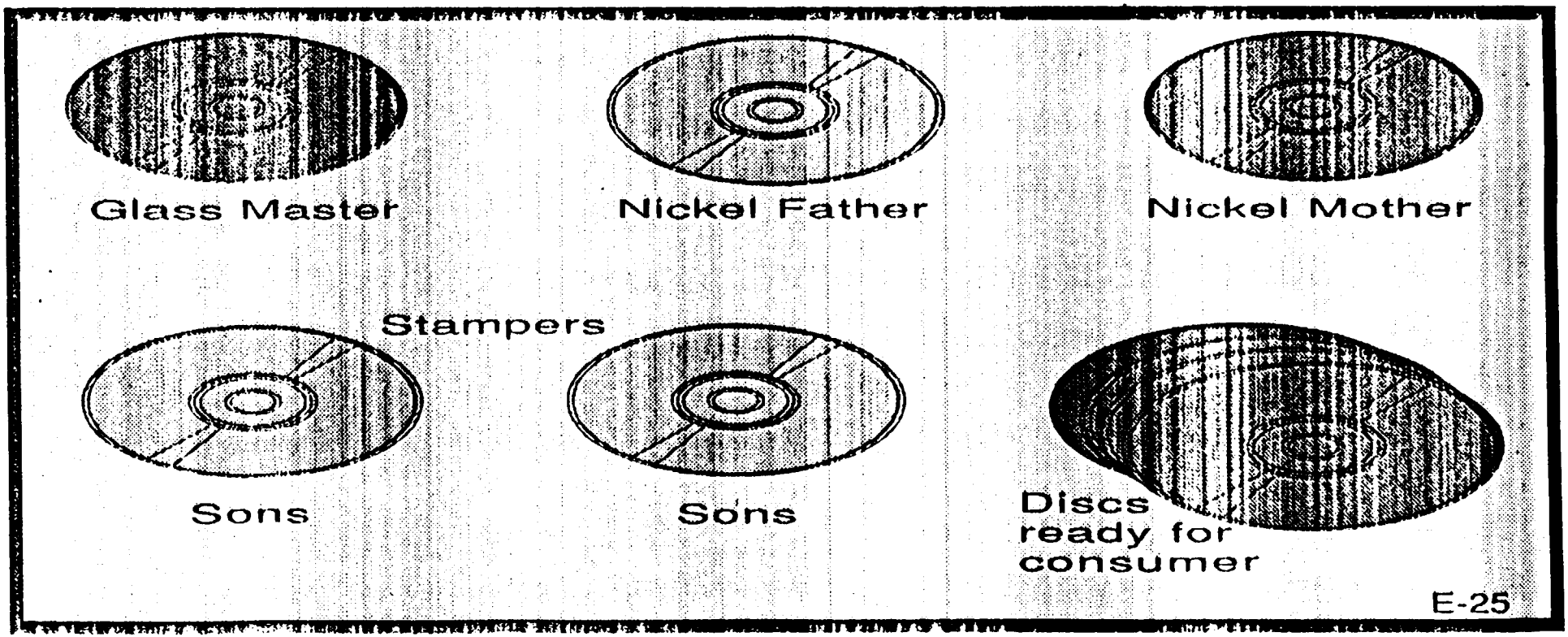
- Before writing the tape(s) for forwarding the mastering facility, search and retrieval testing should be performed to insure functionality
- Simulation software that emulates the performance of the software as if it were on a CD-ROM is available for testing
- The write-once or one-off discs are valuable for beta testing and should be considered since search time is slower on a disc than from the standard hard disk on a computer
- The files are written to an image tape in the exact form in which they will be written to the disc. This form should be in accordance with the International Standards Organization reference ISO 9660
- The disc manufacturer can format the data in ISO 9660 for a fee if required

PREMASTERING, MASTERING, REPLICATION, PACKAGING

- The mastering facility describes making a disc in three steps:
  Premastering
  Mastering
  Replication
- The tape sent to the replicator must be written to their standards so they can read from it
- The 9660 standard data is combined with error correction codes and the necessary data for sync and address are added
- The normal 8-bit byte is expanded to 14 bits per word to permit the laser reader to read the pattern
- The data is then etched to a master disc made of glass and coated with a light-sensitive material, this process is similar to exposing film

PREMASTERING, MASTERING, REPLICATION, PACKAGING--Continued

- Disc replication is achieved as shown in the inset below:
- Glass to nickel, to nickel again, then to plastic stampers



- Packaging options: folders, tea bags, and more, but the most compatible with production are the jewel boxes

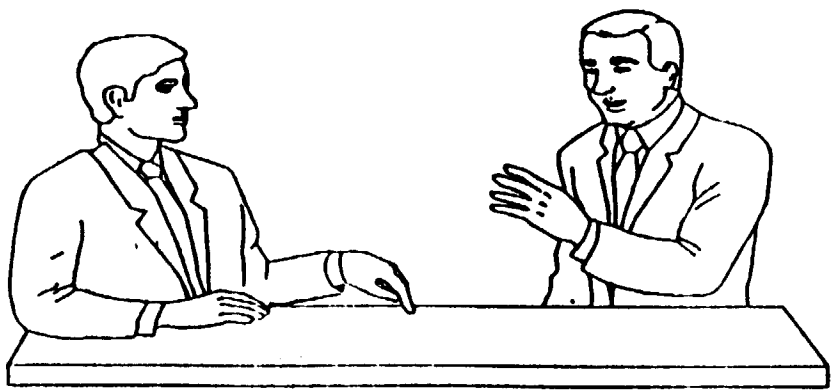
WRITE-ONCE COMPACT DISCS

- Data to disc CD-ROM publishing is in operation at GPO
- This system provides the capability to cut a disc before mastering as a test or may be used for 1 of a kind production
- Write-once discs are NOT "WORM" discs. The write-once type does comply with ISO 9660 and 10149 standards
- Write-once discs have the same 640Mb capacity and the error correction retrieval software are the same as those manufactured at CD disc replicators
- CD-ROM applications and databases can be verified on the write-once disc before mass production
- Write-once discs cannot be overwritten, they are not as long-lasting as factory made discs, shelf life is probably seven years as opposed to 25 years or more with the others

How to get started !

call or meet with your . . .
Departmental Account Representative
or someone in the Electronic
Publishing Management Section

and . . .



arrange a meeting !

Be prepared to discuss some of these specifics about your publication . . .

1. An overview of the product requirements
2. The format of the information base
3. Discussion of software packages
4. Search detail requirements
5. Look at samples of data and hard copy

. . . filling out the questionnaire in your folder prior to a meeting may help

Now It's Our Turn ! ! !

GPO will contact you with:

1. Data analysis prepared by Graphic Systems Development Division (GSDD)

2. Estimates of time to prepare data as CD-ROM information source

When we get the "go-ahead" we will:

1. Use the software build engine to prepare data for writing to an image tape

2. Create a write-once disc for beta testing if requested or forward image tape to manufacturer

Attachment 5

Authoring Software Vendors

Authoring Software

Title	Platform	Vendor	Comment	List Price
Authorware Professional	MS-Windows Macintosh	Macormedia (415)252-2000	Authoring tool for interactive learning	$4995
Director		Macormedia (415)252-2000	Multimedia production authoring tool.	$1195
Super Show & Tell (SST)	MS-Windows	Ask Me Multimedia , Inc. (800)275-8311	Multimedia slide show authoring tool	$200
INSIGHT Into Multimedia, INSITE into Information, INSITE Lite	MS-Windows	ENIGMA Information Retrieval Systems, Inc. (212)599-1400	Comprehensive authoring tools for complete in-house full-text multimedia production. Builds compound document databases capable of integrating text, video, audio, and animation.	
HyperWriter Professional	DOS MS-Windows	NTERGRAID (203)783-1280	Hypertext linking and full text searching capability	$3995

Attachment 6

Premastering Software Vendors

C-2

Premastering Software

Vendor	Platform	Comment	List Price
CD-ROM Strategies, Inc. (714)453-1702	Windows		$250
Corel Corp. (613)728-8200	Windows OS/2	Inexpensive	$249
Dataware Technologies, Inc. (617)621-0820	Windows Sun OS	Includes CD-recorder hardware	$2,000-7,000
Electroson USA (215)617-0850	Windows Macintosh Sun OS		$249(multi-media license $799 for Windows /Mac) $1,695 (Sun)
incatsystems (408)379-2400	Windows Macintosh		$1,495
OMI International (408)376-3511	Windows Macintosh		

Attachment 7

Indexing and/or Database Software Vendors

Indexing and/or Database Software

Title	Platform	Vendor	Comment	List Price
TextWare	DOS MS-Windows Macintosh	Textware Corp. (801)645-9600	Includes full text indexing and searching, hypertext linking, multimedia capability.	$695 - $2995
Filio Views - Infobase Production Kit	DOS	Filio Corp. (801)344-3648	Includes tools to build information databases with linkage to source document. Imports files from major word processing packages.	$895
Report Tracer	DOS MS-Windows	Trace (408)441-8040	A PC-based index and retrieval software package that automates computer report storage and retrieval using CD-ROM. Designed for storage of computer output in ASCII form, it is capable of storing 500,000 pages on a single CD-ROM. Capable of outputing upto 800 pages a minute.	$10,000
Alchemy for CD-R	Windows	Information Management Research Inc.	Provides tools for indexing, data management, searching, and retrieval. Database searches via full text, key word, query-by-example (QBE), and table of contents.	$995

Attachment 8

One-Off CD Shop Price Schedule

Compact Disc Mass Replication Service

The One-Off CD Shop Washington is ready to service your CD-ROM mass replication needs. Mass replication should be ordered for runs of more than thirty discs; for lesser runs one-off discs will be faster and more cost-effective.

Replication Prices

Replication involves the following steps. The prices are FOB Washington, are subject to change and are for estimation purposes only. Please call for a custom quote for your specific job.

Mastering: We will record a premaster disc, it will be sent to you for verification before it is ordered from our replication facility. See One-Off CD-ROM pricing.

CD-ROM Mastering and Replication: The following charges for mastering and replication of your CD includes assembling discs with your insert materials in a shrink-wrapped jewelbox.

Quantity	Total	Per Disc
100	$1035.50	$10.36
250	$1337.50	$5.35
500	$1550.00	$3.10
1000	$2125.00	$2.13
2500	$4325.00	$1.73
5000	$7675.00	$1.54

Expediting Your Service: Standard turnaround for mass replication service is 15 business days.

You can secure faster service at the following rates:

6-11 days	$250.00
2-5 days	$700.00

Additional Services

Disc serialization:	$.20 per disc
Offset disc printing:	Call for details
Printing inserts and booklets:	Call for details

Mass Replication Schedule

To ensure coordination of the mass replication process, we ask that you adhere to the following schedule:

One week in advance of disc replication:

Deliver your data, any printed material to be inserted in the disc packaging, and your disc art on film positives to the One-Off CD Shop.

Five days in advance of replication:

A premaster check disc will be forwarded to you for approval.

Three days in advance of replication:

Return your check disc to us for any changes or for mastering. Changes due to alteration in your data will result in an additional charge.

Mass Replication Printing Specifications

Please call for complete booklet, back panel and disc imprint printing specifications. We can provide you with electronic templates for all CD artwork.

Prices effective August 22, 1994. Subject to change without notice.

THE One-Off CD SHOP WASHINGTON, INC.

Compact Disc Custom Recording and Mass Replication Service

4590 Graphics Drive
White Plains, MD 20695
Phone: (301) 870-0480
Fax: (301) 843-6339
Compuserve: GO ONEOFF

One-Off Compact Disc Production Service

The One-Off CD Shop Washington produces custom (one-off) compact discs from your data in quantities as low as one. One-off service is ideal for archiving, testing, premastering, and small run distribution.

All one-off discs include our standard three business day turnaround, customized disc labeling and a jewelbox with a customized front panel. Call for custom quotes on jobs involving a large projected volume or repeated service.

One-Off CD-ROM Disc Prices

The following information and pricing are valid for CD-ROM discs.

First Copy (Premaster):

Up to 150 MB*	$115.00
151-650 MB*	$190.00

Copies of the Premaster:

1-5	$75 each
6-10	$70 each
11-20	$60 each
21-30	$50 each

Indexing: Please call for details.

Structuring: Arrangement of the directory structure of your CD, and layout of Macintosh Finder windows to your specifications is done free of charge for up to ten minutes. Subsequent structuring is performed at an hourly rate of $75.

*Data size is determined by image file size for ISO 9660 and by

System 6 Compatibility: System 7 compatibility is standard for Macintosh discs, but we can ensure System 6 compatibility for $18.85 per title.

Verification: While we spot check all discs, we can only guarantee our CD-ROMs if we have performed bit-by-bit verification. Verification charges start at $38.00 per disc.

Additional Services

Expediting your Order:

2 business day turnaround:	$115.00
1 business day turnaround:	$195.00

There is an additional $75 charge for weekend service.

Fulfillment: We can keep an image of your CD-ROM on site and quickly create duplicate discs and distribute them to your clients. Please call for details on our fulfillment service.

One-Off CD-I Discs: Image-to Disc and Script-to-Disc services are available.

One-Off Audio Discs: Red Book audio services are available.

Getting your Data to us

Bring us your hard drives, tapes, or cartridges. We currently accept Floppies, QIC 40/80, 8mm Exabyte, 4mm DAT, Bernoulli, Syquest, and SCSI hard drives. Please call for other formats, or for a one-day SCSI hard drive loan. The first unit (ie. cartridge, tape, drive) of any data transfer is free of charge. On-site data transfer or disc replication can also be arranged.

Supplies

The following is a partial list of CD supplies. Please call us if an item is not listed, or if you would like us to send you samples.

CD-ROM Tricaddies: A Tricaddy is both a jewel case and a caddy, with a convenient compartment for an insert, right on the caddy.

1-9	$4.99
10 or more	$3.99

CD-ROM Packaging: A few of the many options of Univenture packaging are listed below. Please call us for samples. Prices below are for quantities under 500. A variety of colors are also available.

Viewpak	$.25
Jewelpak binder page	$.38
Imagepak	$1.20

CD-R Label Sheets: CD laser labels for recordable discs are available in several colors, 8 labels per sheet. Call us for more information about our custom printed labels.

50	$19.00
100	$30.00
250	$75.00
500	$130.00

Recordable Media: Our media has been tested, and we only sell brands that meet our strict quality standards. Prices below are for quantities under 50 discs.

CD-R 63	$17.00
CD-R 74	$18.00
CD-R 74 (4x)	$17.00

CD Recorders

We carry recorders and premastering software for the power user, as well as the casual CD-ROM producer. Some systems are suitable for archiving, while others allow you to create hybrid, mixed mode, and audio discs, right from your desktop. We use these systems every day, and we can help you make the choice that is best for you. Product specification sheets on the latest equipment are available by request.

Attachment 9

Duplicating Services

Duplicating Services

3M Optical
1425 Parkway Dr.
Menomonie, WI 54751
Telephone: 715-235-5567, 715-235-2220; Fax: 715-235-4608

American Helix/KAO
1857 Colonial Village
Lancaster, PA 17601
Telephone: 717-392-7840; Fax. 717-392-7897

America Disc., Inc.
2525 Canadian St.
Drummondville
Quebec, Canada J2B 8A9
Telephone: 819-474-2655; Fax: 819-478-4575

BQC
2121 South 35th St.
Council Bluffs, IA 51501
Telephone: 712-328-8060; Fax: 712-328-0490

Cinram, Inc.
1600 Rich Rd.
Richmond, IN 47374
Telephone: 371-962-9511; Fax: 317-962-1399

Cinram, Ltd.
2255 Markham Rd.
Ontario, Canada M1B 2W3
Telephone: 416-298-8190

Digital Audio Disc
1800 N. Fruitridge
Terra Haute, IN 47804
Telephone: 812-462-8100; Fax: 812-466-9125

Disc Mfg., Inc.
3500W. Olive Ave.
Burbank, CA 91505
Telephone: 818-953-7790; Fax: 811-953-7791

Micro Electronic Products, Inc.

3621 Westwind Blvd.
Santa Rosa, CA 95404
Telephone: 800-736-0269; Fax: 707-576-7704

Sanyo Laser Products
1767 Sheridan St.
Richmond, IN 47374
Telephone: 317-935-7574; Fax: 317-935-7570

Attachment 10

Periodicals

Periodicals

CD-I World
49 Bayview
Suite 200
PO. Box 1358
Camden, Maine 04843
Telephone: 207-236-8524; Fax: 207-236-6452
Ten issues/year; subscription $37.50 US, $78 Mexico and Canada, $146 elsewhere.

The CD-ROM Directory
TFPL Publishing
22 Peter's Lane
London EC1M 6DS
United Kingdom
Fax: 44-71-251-8318
Telephone: 44-71-251-5522
European publisher:
UniDisc
3941 Cherryvale Ave
Soquel, CA 95073
Telephone: 1-408-464-0707; Fax: 408-464-0187
American Publisher:
CD-ROM Librarian
Meciler Corporation
11 Ferry Lane West
Westport, CT 06880
Telephone: 1-203-226-6967

CD-ROM Professional
Pemberton Press, Inc.
462 Danbury Road
Wilton, CT 06897-2126
The magazine for CD-ROM publishers and users
Bi-monthly; subscription $86 US and Canada, $121 elsewhere.

CD-ROM Professional Inside News
Pemberton Press, Inc.
462 Danbury Road
Wilton, CT 06897-2126
Monthly; subscription $345; includes fax Broadcast news flashes.

Digital Media
A Seybold report
Seybold Publications, Inc.
Box 644
Media, PA 19063
Telephone: 1-215-565-2480; Fax: 1-215-565-4659
Monthly; subscriptions $359 US, $401 Canada, $413 foreign.

Envisioneering
Kyra Communications
3864 Bayberry Lane
Seaford, NY 11783
Telephone: 1-516-783-6244, Fax: 1-516-679-8167
Tracking multimedia technologies driving tomorrow's markets
24 issues/year, subscriptions #595 corporate, $395 domestic. International delivery add $200

Information World Review
Learned Information (Europe) Ltd.
Woodside
Hinksey Hill Oxford OX1 5AU, U.K.
Telephone: 44-(0)-865-7302275; Fax: 44-(0)-865-736354
The information community newspaper
11 issues/year, subscriptions L32 per year.

Inside CD-I
The official news magazine of the CD-I Association of North America
11050 Santa Monica Blvd
Los Angeles, CA 90025
Telephone: 1-310-444-6519; Fax: 1-310-478-4810
Quarterly: subscription cost for non-members $85 or $25/issue

The Interactive Exchange
Monitor Information Services
Future Systems, Inc.
PO Box 26
Falls Church, VA 22040-0026
Telephone: 1-703-241-1799: Fax: 1-703-532-0529

The marketplace for the interactive professional
Monthly; free to subscribers to Multimedia &Videodisc Monitor and members of the International Interactive Communications Society and Interactive Multimedia Association. Others $15 annually.

NewMedia
HyperMedia Communications, Inc.
901 Mariner's Island Blvd.
Suite 365
San Mateo, CA 94404
Telephone: 415-573-5170; Fax. 415-573-5131
Subscription inquiries to:
Customer Service Dept.
NewMedia PO. Box 1771
Riverton, NJ 08077-9771
Telephone: 609-764-1846
Multimedia technologies for desktop computer users
Monthly; free to "qualified new media professionals," otherwise $48 US, $82 Canada/Mexico, $96 elsewhere.

New Media News
The Boston Computer Society
One Kendall Square
Cambridge, MA 02139-1562
The Boston Computer Society hypermedia/optical disk publishing special interest group
Quarterly; subscription to members only; subscription to two BCS publications and other privileges included in cost of membership. Various membership plans available, including international.

MPC World
524 Second St.
San Francisco, CA 941077
Telephone: 415-267-1755; 800-274-2815 (subscriptions); Fax: 415-281-3915
Bi-monthly; "charter rate" subscription $14.95, newsstand price $3.95/issue

Multimedia &Videodisc Monitor
P.O. Box 26
Falls Church, VA 22040 Telephone: 703-241-1799
Covering application, innovation, and technology with interactive video, multimedia, and related fields.
Monthly, by subscription only. $347, + $30 outside the US, Canada, and Mexico

Multimedia/CD Publisher
Meckler Corporation
11 Ferry Lane West
Westport, CT 06880
Telephone: 203-226-6967
The publisher's guide to the multimedia business 11 issues/year, subscriptions $147, $97 individual to home address.

Multimedia Review
Meckler Corporation
11 Ferry Lane West
Westport, CT 06880
Telephone: 203-226-6967
The magazine for multimedia publishers, quarterly; subscriptions $97, $35 individual to home address.

Nautilus
7001 Discovery Blvd
Dublin, OH 43016-8066
Telephone: 800-637-3472 or 614-776-3150
Monthly periodical mostly about CD-ROM in Macintosh and PC (DOS) versions. $ 119.40 per year in US plus shipping and handling.

The QuickTime Forum
WayWithWorks
1455 Cedar Oak Rd.
Placerville, CA 95667
Telephone: 916-621-0468
12-page newsletter, 10 issues/year. Subscriptions $75, back issues $10.

Attachment 11

Glossary

CD-ROM PUBLICATION

Glossary of Terms

Information in this booklet has been copied with the consent of (c)Knowledge Access International, Inc., from their manual entitled "Publish on Disk!". Nimbus Records, Inc. also contributed through the use of their manual entitled "The Road to CD-ROM" by giving several of us an extensive tour of their plant and by donating CD-ROMs, brochures, and instructional materials for presentation.

Philips and Dupont Optical also contributed CD-ROMs and brochures.

access time: how long it takes to retrieve a piece of information

address: code that specifies where items are located on a disc or memory system

algorithm: repeatable procedure for performing a task in a computer such as searching, indexing or sorting data

alphanumeric: letters and symbols used to form a unique instruction

analog: a continuous, variable signal, as opposed to digital which is a series of ones and zeroes

application program: a computer program designed to implement a task such as organizing a data base

application software: program designed to perform a task such as search and retrieve

ASCII: a seven bit code that represents numbers, letters, and control characters (American Standard Code for Information Interchange)

authoring system: software that creates an environment in which the user interacts with the computer to perform tasks

baud rate: the number of bits per second transmitted over a communications connection

blister pack: clear plastic shell packaging that displays the disc in its jewel box

bit: the smallest segment of data seen by the computer

bit mapped image: text or drawings interpreted as a whole picture, a rectangular pattern of numbers that represent a picture

block: a single unit of data

block error correction: a string of bits inserted in each block to ensure that errors can be corrected and data can be recovered

board: short for circuit board, used to control a necessary function of the computer or a peripheral

Boolean search: named for a 19th century British mathematician, uses AND, OR, and NOT as search concepts

bridge program: software that converts one file format to another for transfer between programs

browsing: thumbing through a data base for what rather than where

build engine: the part of a retrieval package that creates indexes and data structures for searching

byte: a sequence of bits, generally eight bits long

caddy: plastic container used to insert disc into disc player

CD (Compact Disc): orginally made for music and standardized by the Red Book

CAV: Constant Angular Velocity, process where the disc rotates at a constant speed for reading

CD-DA: the music disc which accounts or 90 percent of discs produced by manufacturers

CD-I (Compact Disc Interactive): contains prerecorded digital video, audio, and optical text data standardized by the Green Book and used extensively at home

CD-PROM: Compact Disc Programmable Read-Only Memory

CD-ROM: Compact Disc Read-Only Memory standardized by the Yellow Book

CD-ROM disc player: standard type of laser disc player used to play CD-ROM discs

central processor: the part of the computer that controls data transfer by executing instructions from the system's memory

cga: color graphics adaptor, see: ega, mvcga, mvga, msvcga, svga, vga

comma-delimited format: field values are separated by commas, and alphanumeric values are enclosed in quotes and used for structured data bases

compression: a process used to reduce the size of a data file, also reduces the amount of disc space used to store the data and reduces the time needed to download the data to a platform for viewing

CLV (constant linear velocity): disc drive that rotates at varying speeds to allow data to move past the optical head at the same speed

configuration: physical components that make up an electronic system

controller: circuit board that interfaces the computer with CD-ROM player or other device

data base: collection of machine-readable information, sometimes structured in fields within records or grouped in blocks

data preparation: gathering, converting, organizing, and editing prior to indexing, premastering, and mastering

data transfer rate: 150 Kb per second is the standard, which requires more than 1 hour to complete. Newer double speed drives have a transfer rate of 300 Kb but require special disc formatting to benefit from this higher speed

delivery system: the user's computer/CD-ROM system for running the application system

digital: Coding of signals commonly in binary (2) levels designated by 0 and 1. A bit is 1 or 0; a byte is 8 bits; a kilobyte is 1024 bytes; and a megabyte is 1024 kilobytes.

digitizing: process of converting text, pictures, or sound to digital codes, mostly used by Optical Character Recognition to convert typewritten manuscript for computer use

disc: optical disc

disk: magnetic disk

disk operating system: a software program that instructs a computer how to transfer information to and from peripheral devices, CD-ROM Extensions are one example

document: small sections into which large blocks of electronically-stored text are divided and indexed on CD-ROM

DVI: Digital Video Interactive; a multimedia compression technology

ECC: Error Correction Code; used to detect erroneously stored data and changed them to correct value

EDC: Error Detection Code; technique used to detect errors during retrieval

ega: extended graphics adaptor

Eight to Fourteen Modulation (EFM): eight-bit data is expanded to 14-bit data for efficient storage on a CD-ROM

encode: to convert information to machine-readable format

error correction codes: during pre-mastering, an error detection code and an error correction code is added to each physical block of data (2048 bytes) to correct errors when the disc is read incorrectly during retrieval process

field: a category of information in data base

file: a single logical set of data

file system: a logical way to organize data on a CD-ROM so that an application program need not be concerned with the physical location or structure of the data

file inversion: inverted file is an index of every keyword in the data which has been re-sorted into alphanumeric order to speed up the search process

fixed length: records or fields that always occupy the same amount of bytes

flag: sometimes called a field tag, used to mark or identify document structure, and display, put into the data at the time of preparation prior to pre-mastering

formatting: blocking data to logical sectors and blocks and adding information used for retrieval programs

formatter: optional part of a data base retrieval package, used to take raw data and place it in a format for indexing

frame: one complete video picture

full inversion: a means of making an index that includes all meaningful words in a document

full motion video: video reproduction at 30 frames per second

full text: data that consists of words or numbers that are contained in a document (not broken into fields)

gigabyte: approximately 1000 megabytes

global search: searching through the entire data file

graphics: usually scanned images

graphics formats: format for computer storage which includes: PIC, TIFF, GIF, IMG, PCX and others

Green Book: CD Interactive standards book

hard disk: usually refers to workstation resident disk used for loading data and programs for processing

header field: in CD-ROM a segment of the sector set aside for address of the sector

High Sierra Group: refers to standard, also called proposal for logical formatting of files on the CD-ROM

hit: when a search request is successful in searched data

HyperCard: a Macintosh program used to interface with CD-ROM

Hypermedia: an extension of Hypertext that incorporates video, audio, animations, etc.

Hypertext: coined by Ted Nelson in 1965 describes the ability to navigate through text in non-sequential manner using links between words

Icon: a pictorial representation of a function or feature in a computer software program

Image file: contains a bit-mapped image usually linked to a search word in the data source file

Indexing: in CD-ROM is used to locate records or words within a file

Interactive: software technology that permits users to navigate through a subject in a random access fashion

Interleaving: a method for blending data streams such as graphics and sound tracks to achieve simultaneous playback, used on CD-I and CD-ROM XA platforms

Injection molding: squirting molten plastic into a mold

interface: a link between two systems

ISO: International Standards Organization

ISO 9660: international standard for formatting files on CD-ROM discs, based on High Sierra Group Proposal

ISO 646: a standard for character set, ASCII complies with this standard

inverted file: index created through building search method where all words except stop words are placed in alphanumeric order with a pointer to location on disc

JPEG: Joint Photographers Expert Group, designed a set of algorithms to compress still images

kiosk: a self-contained, free-standing, interactive system

keyword search: all words indexed to show location can be found with this search, in fielded data a name found in the name field would be a keyword

kilobyte approximately 1,000 bytes

lan: local area network

land: reflective area between two nonreflective pits

Laser card: the size of a credit card, contains from 2 to 10 megabytes which is approximately 800 pages of digital data

laser pickup: the optical read head used with CD-ROM

logical format: refers to file format or organization of files on a disc

magnetic disk: floppy disk and hard disk with data written on surface using magnetic impulses

magnetic tape: tape used to transport data to mastering facility

mastering: etching the original glass from the data on the premastering tape, this disc is called the master disc

megabyte: 1 million bytes (1,000 kilobytes)

mother disc: used to describe a step in creating a stamping disc, the first step is to pour nickel over the glass master and separate the metal after it hardens, the nickel disc is the mother

MPC: Multimedia Personal Computer is a standard devised by Microsoft which describes a 386 or higher CPU running Windows 3.1 or higher with a sound card and a CD player

MPEG: Motion Pictures Experts Group standards for delivering data at a decompressed rate of 30 frames per second

MS-DOS Extensions: used to overcome 32 megabyte limit when attempting to read a CD-ROM

Multimedia: Converging hardware and software technologies that enable text, video, sound, and images to be presented in a cohesive self-directed presentation

mvcga: mono video color graphics adaptor

mvga: mono video graphics adaptor

msvcga: mono super video color graphics adaptor

Navigation: a word describing the ability to move around in a hypertext application

NISO: National Information Standards Organization establishes standards for libraries, publishing, CD-ROM, etc.

OCR: Optical Character Recognition scans an image of text into an ASCII file for storage on disc

optical disc: a high density storage device such as CD-ROM, CD-I, WORM, and Laser cards all of which are non-magnetic

overhead: indexes used for searching can amount to 40 to 50 percent of the source data causing 100 megabytes to increase to 150 megabytes when data and indexes are combined

parallel: eight bits are sent at a time rather than one at a time which is serial, used to describe a port and how it transmits

PD-ROM: a disc containing public domain software released 2 times a year

PhotoCD: a new disc created by Philips and Kodak for the storage and viewing of photos, CD-ROM XA drives can read single session recordings, but special drives are required to read multisession discs

peripheral: devices that connect to a computer but remain physically free from it

photoresist: chemical coating placed on the glass master disc and will be burned off with laser light

phrase search: groups of words searched individually or together

physical format: the size of a disc and the sector and block construction

pits: holes in the track that do not reflect light

pixel: a numeric code repesenting a point on an image to be scanned, sometimes 300 points per inch and similar to a screen

port: a computer workstation and a device are connected through a port in serial or parallel

polycarbonate: plastic used to create base for discs

premastering: sometimes begins with creating the image tape used for mastering, usually refers to placing data, address blocks, and error correction codes in a combined file and writing to a master disc

proximity word search: allows the user to search for words within or not within a certain amount of words in either direction and is more restrictive than Boolean searches

RAM: allows data to be written or read any number of times

random access memory: RAM, a method of random storage using memory circuits to store data in a non-sequential fashion, can be read and written

read-only memory: a memory device that once written to may not be rewritten to

real-time: computer that gives immediate responses

record: a complete entry consisting of 1 or more fields

Red Book: standard for CD audio

replication: process of making copies of a CD disc

retrieval engine: software that provides access to data stored on CD disc

RGB: Red, Green, Blue, 3 signals computers use to create color

ROM: Read-Only Memory is a disc or chip from which data may be read

scanning: converting a printed page to a digital file, digitizing an image

search key: way of referring to a search field

search hit: successfully finding a term, phrase, etc.

sector: smallest addressable unit of a disc's track

seek: describes the mechanical movement of the optical head when moving across the disc to find a sector; CD-ROM's seek by minute: second: sector address

SGML: Standard Generalized Markup Language, used for marking text for a variety of purposes, ISO standard 8879

sideways search: secondary searches that will return to point of origin when terminated

software: programming and documentation, or methodology used in computer applications

stopword: words that are purposely left out of the indexes because they are not useful in finding information

stamper: can be metal master or discs made from master and used to create transparent plastic CD disc

storage capacity: 680 megabytes equals 220,000 pages of typed copy, 5000 frames of video pictures, 72 minutes of sound, all figures are increasing with improvements in technology

substrate: base material used for CD-ROM disc, stamped with information during molding process

svga: super video graphics adaptor

term: a word zone, group of letters and numbers, or sometimes an important word to search

text file: a file composed of any of the 256 ASCII characters (including upper ASCII)

track: a single data format on a continuous section

tracking: the ability of the laser read head to stay in line with the spiral of pits as the disc rotates

TIFF: Tag Image File Format, used to transfer images from program to program

upper ASCII: characters above 127 and ending at 255

volume: usually refers to an individual CD-ROM disc

vga: video graphics adaptor

wildcard search: searching in word and variations by substituting a symbol in the search string

workstation: the platform with computer and various peripherals such as screen, printer, and modem

WORM: Write Once, Read Many disc, usually not in compliance with the ISO standard

Yellow Book: the CD-ROM standard as defined by Philips and Sony

zone: data separated into specific classes sometimes delimited with tags

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(leave blank)*	2. REPORT DATE June 1995	3. REPORT TYPE AND DATES COVERED NASA Contractor Report June 1995

4. TITLE AND SUBTITLE CD-ROM Preparation: An Overview and Guide

5. FUNDING NUMBERS

6. AUTHOR(S) Ralph E. Daniel, Mark W. Jeschke, James A. Schroer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

RP-198756

11. SUPPLEMENTARY NOTES

Technical Monitor: Terese Ohnsorg, Scientific and Technical Information Office

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Unclassified / Unlimited
Subject category - 61

12b. DISTRIBUTION CODE

13. ABSTRACT

A primer on the options and procedures involved in producing CD-ROM products in a small to medium sized business operation is presented in language that persons with a minimal technical background can easily understand. The capabilities, limitations, and standards of CD-ROM technology are surveyed. Emphasis is placed on CD-ROM production, especially upon design, data conversion to an electronic medium, data file preparation, the use of vendors, and the steps for in-house production of CD-ROM products.

14. SUBJECT TERMS CD-ROM, information systems, information management, production management, quality control, data storage, optical disks

15. NUMBER OF PAGES
130

16. PRICE CODE A 07

17. SECURITY CLASSIFICATION OF REPORT Unclass.	18. SECURITY CLASSIFICATION OF THIS PAGE Unclass.	19. SECURITY CLASSIFICATION OF ABSTRACT Unclass.	20. LIMITATION OF ABSTRACT Unlimited

Available from NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
(301) 621-0390