Summer February 1, 1990

ARRB Data Resource CDRom [including many Australian Household interview surveys] (Unpublished)

Marcus R Wigan

Discussion Note DN 2010

CDRom-1/ISO 9660

THE ARRB INFORMATION RESOURCE CD-Rom

by

M R Wigan

NB This CDRom was withdrawn before production by instruction of the Executive Director as '*Cutting edge issues were no longer to be a primary focus*'

Australian Road Research Board
500 Burwood Highway
Vermont, Victoria
AUSTRALIA 3131

DRAFT Rev 5 2 Feb

February 1990

CONTENTS

APPENDICES

INTRODUCTION

This CD-Rom is the first to be produced by the Australian Road Research Board, and is designed to provide an introduction to the use of this new archival medium, and to ensure that the whole process of creating a more accessible and more readily supported method of issuing some of the work of ARRB may be bedded down to a routine process as far as possible. It has been created using a format which can be read on a wide variety of computers, and does not contain any embedded programs which would restrict the usage of the CD-Rom to a single type of machine. This has also reduced the cost and complexity of the Rom, although the potential information retrieval performance on any particular machine has of course been severely reduced as a result. The samples of different forms of electronic optical publishing provided are sufficient to enable users and potential users to obtain a sense of the basic characteristics of the medium, and the author welcomes all user views on the directions in which performance should be optimised and whether the potential range of users should be altered as a result.

WHAT IS REQUIRED TO READ THE CONTENTS?

For those without a CD-Rom drive, any CD Audio player can be used to access the audio track (Track 1). Any MsDos or Apple-compatible CD-Rom drive with the standard ISO 9660 drivers installed on your machine can read the data on this CD-Rom. These cost $1-2k. The

information on CD-Rom is still fairly diffuse, but one of the best collections of resources is available on a CD-Rom itself. The editor of CD-Rom End User magazine (Helgerson 1989) has collated the major books, specifications, general introductory articles and a complete index to available CD-Roms on a single disc, which costs less that $100 US at the time of writing. ISO 9660 drivers are readily available for Apple, MsDos, Digital Equipment Corporation and Unix systems. Consequently this CD-Rom appears as a set of hierarchically-arranged folders on an Apple Macintosh with the appropriate CD-Rom drivers installed. The drivers provided with the Apple CD-Rom drive and the NEC CDR75/77 Series have both been tested. The same CD Rom will appear as a set of hierarchically-arranged MsDos directories on any MsDos machine with the Microsoft CD Rom Extension drivers installed. Helgarson (1988) has aso published one of the clearest concise summaries of the technical and practical issues involved in creating a CD-Rom.

The 'High Sierra' format (Lambert and Ropiequet 1986) was initially widely used, but once adjusted and accepted as an ISO Standard, ISO was the obvious choice for ARRB. All of the files included on this CD-Rom can be read by any of the machines listed above, but to be able to read the materials or use the special facilities will require the appropriate software to be running on the machine used to access the CD-Rom. The InMagic Binary files can be used only by an MsDos or Unix machine running a suitable copy of the InMagic or Search-Magic proprietary software. However, raw text files of this information are provided for those without access to such software, or for use

and experiments on machines other than those supported by InMagic. [Tables II, III and Appendix A].

The hypertext systems on the CD-Rom require copies of the OWL International Guide system - or the more limited read-only Guidance application. The latter is provided with every copy of Aldus Pagemaker as the help system for both the MsDos and the Macintosh versions of the program were written in Guide. Two separate hypertext directories are provided: one for MsDos users and one for Apple Macintosh users.

The contents of the two hypertext systems are identical, and they have been converted back and forth between the Mac and MsDos using software written for this purpose by OWL International, the authors of Guide. This will explain why the 'Help' provided under both versions refers to both Microsoft Windows and to the Mac. The same master files were (and continue to be) used to create and work on both versions. A Unix version of Guide is also available, but has not been tested with this CD Rom. Both versions contain a series of papers written by Hadgraft and Wigan (1988, 1989*a*,*b*), which explain the project on improving the transmission of engineering information of which this hypertext document forms a part.

OBJECTIVES OF THE ARRB CD RESOURCE ROM

The objective of the entire process is to turn undocumented raw

DATABASE	FUNCTION	VALUE	SIZE	CONTENT	CLIENTS	FORMAT
RoadLit	Basic text InMagic*	Consultant and library requests	<15 Mb ~25 Mb indexed	Selected articles on a a broad front	Current subscribers International	Basic lines of text InMagic Binary*
InRoads	Basic text InMagic*	Archival stability. Follows the current trends of the OnLine market.	~25 Mb ~85 Mb indexed	Australian content & authorship Holdings Abstracts	OnLine users Libraries	Basic lines of text InMagic Binary*
ARRB Authors Publication Record	Basic text InMagic*	Complete record of all papers authored by ARRB staff.	<13 Mb ~7 Mb indexed	Author, title, bibliographic reference and date	Access to the identifiable contribution of ARRB to the literature	Basic lines of text InMagic Binary*
Hypertext: the Flood Estimation Chapter of Australian Rainfall and Runoff: and related materials	Hypermedia documents. Instruction Evaluation and related technical papers. (Hadgraft and Wigan 1988, 1989*a,b*)	IEAust endorsement First in the field Tool for active evaluation of hypermedia in training	~5 Mb	Flood estimation ILSAX Graphics Excel sheets	IE Aust LGA Water	Guide-specific files, linked to lines of text. Both Mac and MsDos versions of Guide
Data Publishing. Household interview surveys data used in SR38** File formats documented in ATM24***.	Numeric data publishing Archival data, programs, texts	Archival stability. Self documented archive example for field and market review	~200 Mb	Data files, Texts Programs	Archival Educational Planning State Road and Transport Authorities	Lines of text. MsWord MsDos format. This format can be read by both Mac and MsDos.
Full text and graphics archive . Full documents, complete with their text and graphics.	Full text of documents trials for information retrieval	Archival. Experiments Complements the HIS data.	~50Mb	SR38**, and other related text & graphical stock holdings	Archive, libraries, replaces old stock holdings	Lines of text. MsWord MsDos format. This format can be read by both Mac and MsDos.

* *Denotes an indexed file structure* ** *Wigan (1987)* *** *Wigan (1988)*

Table I. THE CONTENTS OF THE ARRB INFORMATION RESOURCE CDROM

numbers into a useful and self contained information source for all parties to evaluate. This implies that the definitions are as follows:

Data	Basic numerical, textual or graphical records
Product	Data plus sufficiently detailed descriptions, specifications and access programs and procedures added to make the data accessible without recourse to information not provided with the data.
Information	Interpreted data, plus explanations, qualifications, limitations and applicability of the findings.

Data is plentiful, but the additional materials required to use it with confidence have always been considerably harder to find. Frequently, once a data source has been located, the ability to actually read the records can be a severe problem. The physical formats on which the data may reside could be printed sheets (which require extra effort to make machine readable for analysis to become economic), magnetic tapes requiring unusual types of tape drive (7-track etc) as used by some of the older mainframes and minicomputers, or without adequate documentation of the data records written on them. The practical importance of this common situation is reviewed in Wigan (1985), as the secondary users of transport and road information are now becoming more widespread.

DATA RECORDED ON THE CDROM

The contents of this initial ARRB CD Rom are listed in Table I. The basic criteria for inclusion were:

1. Demand or potential demand from clients
2. Relevance to active projects.
3. Provision of a complete example of data, information, access methods and explanatory reports.
4. Demonstration of a current or potential ARRB information product or service

PRODUCTS

Few of the data sets on this initial ARRB Resource CD Rom are the complete resource, and the contents of each of the different types of files are detailed below.

For example, the **RoadLit** information covers most of 1989, but the value of RoadLit is in its weekly immediacy, as well as its selective identification of relevant items in the literature. This segment provides a practical example of the type, format and nature of this continuing information resource. There are two forms of RoadLit file: one is a simple text file containing the material, and the other is a fully-prepared binary file set up for the InMagic text retrieval package. The former can be read on any computer which can mount CD Roms written in the ISO 9660 format, the latter can only be used by machines running some version of the InMagic or SearchMagic software. No such searching software is included on this CD Rom.

InRoads is a substantial database, containing abstracts and key-

words in addition to both bibliographical data and the locations of Australian holdings of the documents referenced. The period covered by the files on this CD Rom range up to the end of 1986. The fully uptodate InRoads database is maintained on the AUSTRALIS online service operated by CSIRO. Further details on this and other roads-oriented data bases are given in Price (1989). This file is also provided in both simple text and InMagic binary formats.

InRoads includes all ARRB publications published since 1960, all books etc. catalogued by the ARRB library since January 1984 and some older books which have been recatalogued , such as OECD publications. It also includes documents about Australian roads and road transport published since 1977. Some of these documents are not held at ARRB, but are held in a number of co-operating libraries, such as State Road Authority libraries. Up to the end of 1985 InRoads has IRRD keywords (OECD 1985). From 1986 local InRoads descriptors are used and the record format is slightly different.

The production of future CD-Rom versions of InRoads should be done in conjunction with the other interested library and online services. However, the present ARRB resource CD Rom will enable librarians to consider the likely worth and practical value of such a series, and provide a sound basis for assessing and costing the effort and time required for all but the important issue of CD-Rom layout and performance optimisation to match particular retrieval software.

The **ARRB Authors publication record** is up to date to the end of 1989, and contains all works in the ARRB databases for ARRB authors. It therefore comprises a near-complete record of the published work of the ARRB staff. This file is also provided in both simple text and InMagic binary formats, and the record structure is given in Appendix A.

The **Hypertext: Flood estimation** is written using a product called Guide. This product links together text and other files as well as creating its own special formats. Consequently this section contains free text files, Guide files and even some computer programs, as Guide can also run and control other computer programs from within a Guide document. This ties a single version of a Guide document to a single operating system or computer, and so two different versions are included on the CD-Rom, ready to be executed by Mac or MsDos versions of Guide or Guidance respectively. Both versions also contain hypertext versions of the published papers produced on the ARRB/Monash project involved, and both launch other programs as part of their operation. Microsoft Excel is called for by both versions, although both versions can read the Excel spreadsheet formats written by either.

The **Data Publishing** component of this CD-Rom is the data sets used to carry out the analyses for Wigan (1987). These are selected and reformatted versions of the personal travel records from a number of Australian urban household interview surveys. The fields have only a restricted amount of accessible information, and the

SOFTWARE	OPERATING SYSTEM	WRITTEN BY	READ BY	REQUIRES
Symmetry PictureBase	Mac	Mac	Mac	PictureBaser Retreiver
Microsoft Word	Mac	Mac for MsDos	Mas +MsDos	Word on Mac or MsDos
Microsoft Word	Mac	Mac for Mac	Mac	Word 4 on Mac
Microsoft Excel	Mac	Mac	Mac +MsDos	Excel: Mac or Windows
Microsoft Excel	MsDos +Windows	MsDos	Mac +MsDos	Excel: Mac or Windows
OWL Guide	Mac	Mac	Mac	Mac Guide or Guidance
OWL Guide	MsDos +Windows	MsDos	MsDos	MsDos Guide or Guidance
InMagic	MsDos	MsDos	MsDos	InMagic or SMagic
Text	MsDos	MsDos	Mac +MsDos	Mac or MsDos
Text	Mac	Mac	Mac +MsDos	Mac or MsDos
Sound	None	Direct to CD Rom	Mac +MsDos	Audio track access

Table II. FILE READING REQUIREMENTS AND COMPATIBILITY

(in the MsDos Microsoft Word format which can be accessed by both MsDos and Apple system), the Excel spreadsheets from which most of the diagrams were generated, the Excel picture files generated by the Apple Macintosh version, and the complete PictureBase searchable library of the complete set of diagrams from Wigan (1987).

All of the files on the disc can be located under MsDos or the Apple Macintosh file system, but the contents of the files may not appear to make sense unless the appropriate programs are used to access the files (Table II).

Data is held in basic text file format, and so can be read by almost any application on either type of machine. The reports were written on the Mac, and in some cases included graphics. Although the Mac version of Word can write MsDos format files (as has been done) the graphics are not transferred. Recent release versions of Excel write files that can be written and read on either type of machine. Guide documents have their own internal structure, and require a copy of Guide (or Guidance) on the appropriate machine. InMagic binary files can only be read by InMagic programs, which are available under Unix and MsDos only. The PictureBase library can only be accessed by a copy of PictureBase (or PictureBase Retriever) running on a Mac.

Shortly it will be possible to create a similar resource to the PictureBase libraries and the Guide hypertext documents using Hypercard or Plus on the Mac, and Linkway or Plus under MsDos with Windows.

This is not yet possible, as the necessary products have not yet been widely released. This will materially ease the task of including graphical and hypertext materials, yet keeping the CD-Rom readily accessible to both the major types of machines. As Plus and Guide are both due out under Unix in 1990, this is a development could have a wide and substantial impact.

The sound file is a straight audio track introducing the disk, and discussing the materials on the disc and the objectives of the project. It may be read by either type of machine, using the audio track access programs available for both. This range and variety of formats and access advice should assist users to gain an appreciation of the advantages (and disadvantages) of making various choices for their own archival projects.

PRODUCT	DISTRIBUTED IN AUSTRALIA BY
Format GMBH Plus [Apple Macintosh+Windows*]	Trio Technologies, Black Rock, Vic.,Australia
InMagic 7.1 [IBM PC MsDos]	Triad, Melbourne, Australia
Microsoft Word 5 [IBM PC MsDos]	Microsoft, Frenchs Forest, Australia
Microsoft Excel 2.0 [IBM PC MsDos+Windows]	Microsoft, Frenchs Forest, Australia
Microsoft Word 4 [Apple Macintosh]	Microsoft, Frenchs Forest, Australia
Microsoft Excel 1.0 [Apple Macintosh]	Microsoft, Frenchs Forest, Australia
OWL Guide 2.0 [Apple Macintosh]	Software Suppliers, Brisbane, Australia
OWL 2.0 [IBM PC MsDos+ Windows]	Software Suppliers, Brisbane, Australia
Symmetry PictureBase 1.2.3 [Apple Macintosh]	Infomagic, Frenchs Forest, NSW, Australia

*Denotes not yet released

Table III. USED IN THE CD ROM PRODUCTION

The CD-Rom drive used at ARRB to test this CD-Rom on both Macintosh and IBM PC was a Nippon Electric Corporation NEC CDR-75 self powered SCSI external drive with CD audio stereo outputs.

ACKNOWLEDGEMENTS

The data published on this trial CD-Rom has been extracted from the considerably larger urban transport survey datasets was collected from the different States of Australia. Only a limited amount of the data on this CD Rom has been documented for further use by those with access to a copy of the Rom. We are pleased to acknowledge the permission of the following organisations to include these trial selections from their data.

.	Department of Main Roads, Queensland
.	Department of Roads - Transport Tasmania
.	Ministry of Transport, NSW
√	Ministry of Transport, South Australia
√	National Capital Development Commission, ACT
√	Transit New Zealand
	VicRoads, Victoria

Some of the text used in the Flood Estimation hypertext system that were extracted verbatim from *Australian Rainfall and Runoff* (Pilgrim 1987) has been made available for these research and evaluation

purposes by the Institution of Engineers Australia. Rights to the hypertext structures in which this text is contained are held by the authors, Monash University and ARRB.

The diagrams in the graphics libraries may be freely reproduced, as long as a full reference is made either to the relevant document (as in Wigan 1988), or to this CD-Rom (Wigan 1990).

The production of this CD-Rom required the cooperation of a number of ARRB staff, particularly Vija Walz, who created the InRoads, InMagic, RoadLit and ARRB Author databases in the form required for this publication. The project staff also included Ray Price, the head of the Library services at ARRB when the project was initiated.

DISCLAIMERS

The materials on this CD-Rom are provided for evaluation, investigation and assessment purposes, and no warranty is offerred for the fitness for any other specific purpose. Any views put forward are those of the authors of the various components, and are not necessarily those of the organisations to which they are affiliated.

THE EQUIPMENT USED TO CONSOLIDATE THE DATA

The device used to consolidate the data files was a self-powered SCSI bus-based Ricoh rewritable optical disc drive. This uses double sided

275Mb cartridges, yielding 540Mb capacity. One side was mastered as a Mac HFS files system, the other as a single MsDos 3.30 partition. This simplified the data consolidation process immensely, and was also the method by which the data for the CDRom was provided in a compact and easily transferred format to the premastering service at InfoOne.

REFERENCES

HADGRAFT, R. and WIGAN, M.R. (1988). Assessment and improvement of the codification and delivery of engineering experience. Proc. *Watercomp '89*. National Conference Publication No. 89/8. Institution of Engineers, Australia, Canberra, ACT. pp. 160-164.

_____ (1989*a*). Towards effective field support and retraining for engineering who use *Australian Rainfall and Runoff*. IEAust. Hydrology and Water Resources Symposium. Christchurch, NZ. National Conference Publication 89/19. Institution of Engineers, Australia, Canberra, ACT. pp. 468-472.

_____ (1989*b*). A possible prototype for a new *Australian Rainfall and Runoff*. Hydrology and Water Resources Symposium. Christchurch, NZ. National Conference Publication 89/19. Institution of Engineers, Australia, Canberra, ACT. pp. 481-482.

HELGERSON, L. W. (1988). CD-Rom publishinsg strategies. PC

Tech Journal.(Oct). 8pp.

_____ (1989). The CD-Rom Sourcedisc™. ISO 9660 Format. Diversified Data Resources, Falls Church, Virginia. DDRI, 540Mb.

LAMBERT, S. and ROPEIQUET, S. (Eds.). (1986). CD Rom: the new papyrus. (Microsoft Press; Redmond, WA.).

OCED (1985). International Road Research Documentation (IRRD) Thesaurus, 1985. 2 Vols. OECD Road Research Programme, Paris.

PILGRIM, D. (Ed.) (1987). *Australian rainfall and runoff.* 3rd Edition. Institution of Engineers Australia, ACT.

PRICE, R. (Ed.). (1989). Databases for road engineers. Research Report ARR 159. Australian Road Research Board, Vermont, Victoria. 38p.

WIGAN, M.R. (1985). The secondary use of transport survey data. In: Ampt, E., Richardson, A.J., and Brog, W. (Eds.). *New survey techniques in transport.* VNU Science Press, Netherlands. pp. 131-148.

_____ (1987). *Australian personal travel characteristics.* Special Report SR 38. Australian Road Research Board, Vermont, Victoria. 278p.

_____ (1988). Data and Pictorial Resources for *Australian personal*

_____ (Ed.). (1990). The ARRB information resource CD Rom. CD ROM-1/ISO 9660. Australian Road Research Board, Vermont, Victoria. 450 Mb and 10 min of CD Audio. ***[The present CD Rom]***

APPENDIX A

RECORD STRUCTURES FOR THE TEXT DATABASES

INROADS RECORD STRUCTURE FIELD DESCRIPTION

RC	Record control number
DA	Date (year of publication of document)
LA	Language
DT*	Document type
BL	Bibliographic level (whole or part of document)
RT*	Record type
IC*	Input centre
BA*	Batch
TI	Title
AU	Author (personal name) and affiliation
CA	Corporate author (organization)
CO	Conference details
PU	Publication details (report no, publisher, etc)
AB	Abstract
FI	Field (IRRD broad subject field)
DE	Descriptors (InRoads)
ID	Identifiers (IRRD & free test)
AV	Availability (library)
NU*	Number
NT	Notes

ARRB AUTHORS RECORD STRUCTURE FIELD DESCRIPTION

RC	Record control number
AU	Author (personal name) and affiliation
DA	Date (year of publication of document)
TI	Title
CO	Conference details

ROADLIT RECORD STRUCTURE FIELD DESCRIPTION

RC	Record control number
DA	Date (year of publication of document)
LA	Language
DT*	Document type
BL	Bibliographic level (whole or part of document)
RT*	Record type
IC*	Input centre
BA*	Batch
TI	Title
AU	Author (personal name)
CA	Corporate author (organization)
CO	Conference details
PU	Publication details (journal name etc.) (limited information)
DE	Descriptors (InRoads) (limited indexing)
AV	Availability (ARRB library only)

APPENDIX B

TECHNICAL ISSUES IN CREATING A CD-Rom

A CD-Rom contains 72 to 74 minutes of CD Audio or 660 Mb of data, or a combination of the two. ISO 9660 (ISO 1988) was chosen for this CD-Rom as it is an agreed international standard for which all parties are now writing drivers and mastering their CD-Roms. It contains a carefully defined specification of the exact directory formats to be implemented, and provides some enhancement of speed even for plain CD-Roms with no specific indexing. Drivers written to access such CD-Roms display the directory structure as a series of folders in graphical systems or as a series of nested directories under MsDos or Unix.

Top-level logical structures

The structure of the ISO 9660 directories, which are interpreted by the specialised device drivers on each operating system and computer wishing to access such CD-Roms, is shown in the next figure.

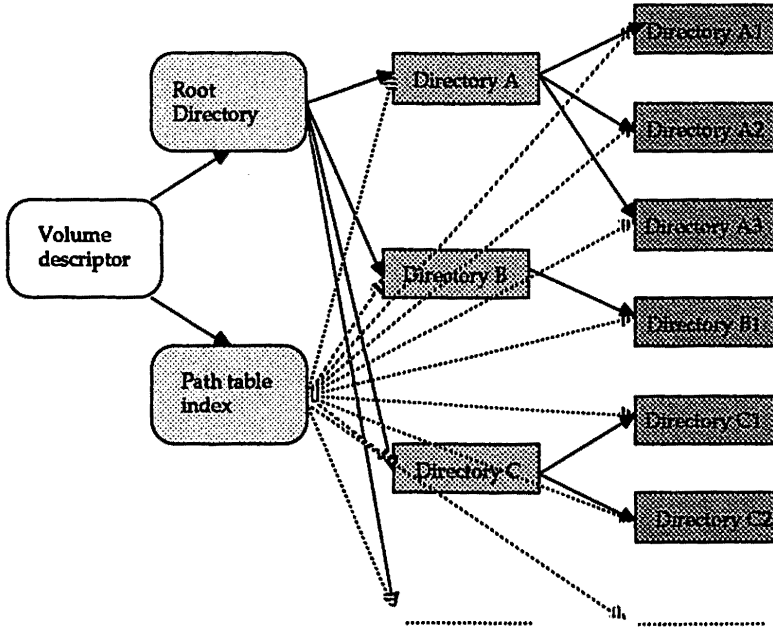
The ISO 9660 Standard has been supported by the appropriate extensions to the operating systems of the IBM PC [MsDos], the Apple Macintosh [Mac Finder o/s], the Apple][[ProDos], DEC [VMS], Sun [Unix], Apollo [Unix] to name but a few. This means that a Mac will show a series of nested folders, an MsDos system a series of nested

directories, and a Unix system a set of nested Unix directories. Sun have also mapped the CD-Rom onto their distributed network file system (NFS) for even wider ranging access. Consequently CD-Roms can be mounted as general network information resources on all the major networks: Novell, Appleshare and Sun-NFS included.

The manner in which the data is written onto the CD-Rom is exactly the same as the audio information is written onto a domestic CD audio disc - but there is no guaruntee that the information in each of the fields will be recognisable by the particular computer you are using - this should be clear from the fact that audio CDs can also be mounted and recognised on many of these CD-Rom drivers.

The major difference between CD audio discs and CD-Roms is that CD-Roms cannot put up with the error rates permissable on a domestic audio disc. Both the CD creation process and the CD-Rom drive itself is of higher quality in terms of the integrity of the data written or read back. Most CD-Rom drives can also be used to play CD audio discs under computer control, and as many CD-Roms may contain sections of CD audio quality sound, these CD-Roms can be played on a standard audio CD player, as long as the appropriate tracks are chosen. It is conventional to place all the data on track 'zero' and the audio data on tracks 1 onwards.

It is not enough to provide a means for different operating systems to find and access the directories and files stored on a CD-Rom. The interpretation of the information held describing these files also



THE DUAL DIRECTORY ACCESS STRUCTURES FOR ISO 9660 CD ROMS

requires some careful specification. The two major microcomputer architectures are the Intel (8088, 8086, 80286, 80386, 80486) and The Motorola (68000, 68020, 68030) families. These two families of chips interpret data held on a linear file in two quite different ways.

- With the least significant byte first (Intel) and;

- With the most significant byte first (Motorola, most mini-computers and mainframes).

Consequently all of the dual directory directory information requiring more than one byte is held on ISO 9660 CD-Roms in both of these two different forms.

The ISO 9660 standard also requires that every directory must contain information on the file location, length, time and date of creation, name of the file, the interleaved structure of the file and a series of flags specifying different aspects of the file or directory - including read protection.

Physical and low-level structures

The basic structure of the CD-Rom is as a series of about 330,000 sectors of 2048 bytes, and applications can of course be written to ignore the ISO directory structures completely. Not surprisingly, the record format specifications for submitting the 6 or 7 full magnetic tapes required to fill a CD-Rom to the production facility is therefore an ANSI format tape using a fixed block length of 2048 bytes.

The hierarchy of information handling on a CD-Rom covers the basic form of writing on the disc (dents or pits in the physical CD, written and read by laser), the error correction structures required to obtain

the very high levels of data retrieval accuracy and reliability reliability required. The standards acceptable for hard discs are far far too low for use on a CD-Rom, and indeed even hard discs are more demanding than CD audio for accuracy in retrieval: every bit (literally) matters on a CD-Rom.

The error correction hierachy for CD-Roms is therefore quite complex, and is a basic piece of the software embedded in the CD-Rom reading device used to extract the information written on the disc. The application of the Reed-Solomon decoding is also part of the task of the CD-Rom interface, and is not the problem of the programmer wishing to use the data on a CD-Rom. CD-Roms generally maintain a data transfer rate of about 150kb/sec: roughly one third of the access speed of a reasonably fast microcomputer hard disc. The data rates defined for the various devices to ensure reliable data retrieval comprise:

CD data rate: 8 bits/byte*24 bytes/frame*98 frames /subcoding block*75 subcoding blocks/second= 1,411,200 bits/sec

CD digital audio rate: 44.1 k samples/sec*16 bits/sample*2 channels=1,411,200 bits/sec, or 175 kb/sec in terms of the equivalent in data space.

It is of course no coincidence that these figures are the same. At this level of detail, the subcoding channels contain extra information that is important to understand. There are 8 subcoding channels, and two of them have a standard usage. Channel P is a track separator to indicate CD audio, and channel Q contains encoded information

on the track number, type and location in terms of both frame number and minute/second codings. These types of information are required and used by CD Audio discs and domestic players. Subchannel Q's control bit fields include specifications of audio and data tracks, but the data tracks **must** come first on the CD-Rom.

The use of CD audio tracks on a CD-Rom requies a great deal of stroage space on the CD-Rom: about 9 Mb for each minute of full stereo CD Audio quality sound. It must also be remembered that a dual mode audio/data CD-Rom drive can play only audio or data at one time: not a combination, as a result of the dedication of the full bandwidth to one or other mode.

The next level up requires computer data structures to be defined. Each CD-Rom data track contains blocks of 2532 data bytes: 12 bytes of synchronisation field information, a 2kb block of user data and the remainder a 4-byte header header field (containing the block address as the minute, second and frame nunber of the block), and if the mode of the block is one, 280 bytes of additional error detection and correction information. The technique adopted permits error bursts of up to 450 bytes to be completely corrected on the fly, plus an additional four error correction bytes to clean up the 24 data bytes. The data stream is then de-interleaved to provide the correct byte order required.

The ISO 9660 logical file structures previously discussed are defined on top of this physical block structure.

Operating system extensions

Apple have defined extensions to the ISO 9660 standard to permit full support of Mac HFS and ProDos file systems (Atsatt et al 1989). Apple][ProDos and GS/OS require a file type and an auxiliary file type, while the Macintosh HFS requires both file and creator types and file attributes. Atsatt et al. specify an exact series of Directory SystemUseField entries to ensure full compatibility, and the following discussion is condensed from this source. The Directory record use field is specified by ISO 9660 as follows:

byte	DirectoryRcdLength
byte	XARlength
struct	ExtentLocation
struct	DataLength
struct	RecordingDateTime
byte	FileFlags
byte	FileUnitSize
byte	InterleaveGapSize
long	VolumeSequenceNum
byte	FileNameLength
char	Filename [FilenameLength]
byte	RecordPad*
char	SystemUse [SystemUseLength]
byte	RecordPad *(if necessary)*

** (only if necessary to make DirectoryRcdLength an even number: if present it must be $00)*

The SystemUseField, if present (as it will be if these extensions are being used) must begin with a signature word, followed by a one-byte length and a one byte SystemUseID, followed by system-specific information. This structure can be repeated multiple times for multiple operating systems, but must the total length of the combined result (i.e. the SystemUseLength) must be an even number of bytes.

It should be noted that CD-Rom-XA discs are a special case, and any signature word that is included should start at byte 14 of the SystemUseField rather than byte zero. The CD-Rom-XA specification is designed to support multimedia under MsDos, and provides for interleaved audio and data. Unfortunately, some of the products using this emerging 'standard' may not yet be as stable and transferable as one might hope. A major CD-Rom retrieval software vendor recently discovered that their XA 'general' materials worked only with Sony CD-Rom drives. This sort of problem is to be expected for some time yet, so the special requirements of CD-Rom-XA discs will not be addressed in the ARRB information CD-Rom. Apple and Sun have not yet addressed the need for XA compatible drivers, and as this CD-Rom is designed to be used a widely as possible, XA has been ignored at this stage.

Efforts should always be spent not to preclude other systems from using CD-Roms effectively. Multiuser operating systems such as VMS and Unix can also benefit from entries in the extension fields, but it is not clear how this conflicts or otherwise with the Mac and ProDos entries. However, ISO 9660 is now actively supported by

Sun, DEC, HP and other vendors: Sony and IBM are expected soon, and few manufacturers are ignoring CD-Rom as an efficient method of delivering bulk materials of software or documentation, and many have already started to release materials internally in these forms.

Apple have not yet linked their AU/X Unix system to the SCSI manager in the Mac operating system, and so the standard Apple drivers cannot be used - but as their AU/X contains the ability to address a Unix file system, it contains a driver which maps the ISO 9660 directory structure onto a Unix file system. This makes the entire disc accessible as a read-only Unix file system, which can be mounted by the user like any other file system. This is precisely how AU/X is now distributed internally within Apple. For US-based companies the cost can be as low as $1500 US for mastering and $1.50 US per disc: for end users the costs are typically 50% or more greater, even in volume- but costs are still declining.

It is desirable that the access to CD-Roms be unified and uniform over the whole disc and over several operating systems. This is still not always possible. Although options are rapidly emerging, the coverage is not yet complete. Once a CD-Rom has been designed to include a special indexed retrieval structure for the data, then the game changes. There are a number of MsDos+Mac products already available (e.g. Quantum Leap, KRS, KAWare and others for example), and yet others that are restricted to one operating system (e.g. TextWare to MsDos). These are indexing systems, and are capable of turning the data into searchable information with rapid access.

It is at this point that it becomes important to simulate the operational speed for the CD-Rom that one has designed, and as CD-Roms can only be read from, and not written to, many programs may have difficulties as they may require the ability to write temporary files on the same directory as they are reading from. Many unpleasant surprises have been encountered by those who were not aware that the programs they wished to use required such write access until they made use of them on a CD-Rom. It is straightforward to test this when using a large (500+ Mb) read/write optical disc to assemble the materials for a CD-Rom, as all that has to be done is to move the physical tabs that disable writing to the disc. This should always be done.

The indexing facilities built into ISO 9660 may or may not be exploited by the software drivers for a particular operating system or for a particular CD-Rom, so the third party software is usually essential. This may be obtained from many sources, and it is important to note that the use of a third party indexing engine does not remove the requirement to optimise the layout of the data and files on the CD-Rom itself. CD-Roms are laid out as one continuous spiral track of digital information on the glass substrate, and not as a series of continuous circular rings of data.

The 12", 8" and 5" video Laservision discs are of two types: CAV (constant angular velocity) and CLV (constant linear velocity). CAV is in the form of such continuous rings, and provides superb static "Still" frame displays, while CLV is the one continuous spiral track

familiar from vinyl audio LPs. CD-Roms are CLV devices. This means that the placing of the file on the CD-Rom governs access speed, irrespective of the efficiency of the indexing system to locate the portions needed to respond to a given query.

Practical implications:

1) Always use ISO 9660 for your CD-Roms.

2) Keep to 31 characters for the file names, and, where at all possible, to the lowest common denominator for file and directory names to ensure the widest accessibility of the data (i.e. the MsDos requirement of an eight character string, a colon and a further three character string - and not more than 8 levels of nested directory). Not keeping to this recommendation will simply limit the range of machines and operating systems which could be used to access your CD-Rom. Mac and Unix systems will generally be fairly tolerant.

3) Ensure that the Apple-defined entries to the directory/file extension fields extensions in the ISO 9660 specifications are used to permit file Types and Creators to be added to Apple files (this will not affect the other uses of the same files by other operating systems).

4) Include an audio track on every disc which at the very at least describes the contents of the CD-Rom. The use of full stereo CD audio quality PCM (pulse code modulation) recording is fully compatible with the CD-Rom data track standards.

5) Check for CD-Rom-XA extensions specifications and any conflicts that might arise: make provision for XA where possible.

The process of mastering a CD-Rom requires that all the data is set up and indexed in the proper formats with all the associated data to go straight onto a CD-Rom. This translation process to a suitable format requires one of several clearly specified formats to be provided to the organisation responsible to pressing the glass discs. Amongst some other variations (aimed mainly at pure CD Audio disc creation), the Australian Disctronics company requires one of the following formats (Disctronics 1989, although others could be handled.

3/4" U-Matic video tapes [essentially for sound information]:

- In PCM-1630 format
- Continuous, unique and upward running SMPTE time code on channel two
- With a log specifying start of pre-gap, start of user data, end of user data and end of post-gap.

ANSI unlabelled magnetic tapes:

- No labels (header or tag)
- Fixed length records of 2048 bytes
- Block sizes equal to this defined record size
- The number of blocks per tape and the tape number, in sequence from zero
- At least one tape mark to follow the end of data on tape
- A log of directory, subdirectory and file structures required on the CD-Rom.

ANSI labelled magnetic tapes:

- ANSI label corresponding to Level 3 of the ANSI X3.27-1978 standard
- Fixed length records of 2048 bytes
- Block sizes equal to a multiple (not exceeding 30) of this defined record size
- The number of blocks per tape and the number of the tape, both in sequence starting from zero
- At least one tape mark to follow the end of data on tape
- A log of the directory, subdirectory and file structures required on the CD-Rom.

These formats are not particularly hard to create, with the exception of the U-Matic tapes. As the simplest form of audio track to add to a CD-Rom is at the full bandwidth of a CD audio disc, it is advisable that the recording is done in a professional studio, and the tape produced at once. An hour should be allowed for the recording of a 10 minute section, to allow for reruns: if a good professional standard is wanted for the final product, one or two hours of PCM recorded tape editing time is very likely to be needed on professional facilities. The hourly charges (typically $150/hour) are comparable to those for large scale machine time for index construction for embedded software systems such as KAWare during the CD-Rom premastering stage.

Typically premastering houses accept data in the format specified above: this means that the finer details of the SystemUseRecord, signatures etc are handled by the equipment that they use. It is

important that CD-Rom clients understand the issues, however, and are able to express precisely what they want on the tapes that go from the premastering house to the disc pressing plant. The previous discussion is designed to provide a succinct summary of the issues to be covered in such a specification. If one does not know what can and should be specified, then unexpected results can come back in the form of a large (and expensive) stack of unusable silver discs.A suitable checklist to cover the issues that need to be resolved is given in this report as an appendix. It is not designed to be complete, but rather to ensure that the premastering house is able to help you to set out a full and complete specification of what you want.

It is not difficult to write a program to write a series of magnetic tapes in the specified formats, and to consolidate the various sources for the files (be they ASCII, MsDos, Mac, Unix or any other format of files that you require on the CD-Rom). The files must be written in a sensible order, and be associated with a clear directory structure description for them to be usable by the disc pressing plant. Any organisation considering creating purely archival data or text discs should also consider creating their own premastering tapes. If they intend to use embedded special purpose text of data retrieval software, and to optimise the disc layout, then this is not necessarily the best path to follow, as this could be done more effectively in many cases by a premastering service.

A simple check that should always be made is to ensure that the CD-Roms are specified in enough detail to be readable on all the target

systems. In the case of Apple, this means checking which Signatures the CD-Rom drivers for your particular CD-Rom will support. The brief discussion above explains why it matters, but see Atsatt and Bechtel (1989) for a clear expression of all the finer technical details. It is prudent to ensure that a copy of the Apple (1989) developers guide is available to you when dealing with these and other questions. It is equally important to realise that the drivers for ISO 9660 and audio control for CD-Roms attached to a Mac are almost invariably Apple's own drivers: the direct links to other CD-Rom drives are handled by the special drivers from NEC and the other manufacturers. However, the final access stage required to decode the various formats for the detailed file structures are handled by the Apple system folder files named 'Foreign file access', 'High Sierra' and 'ISO 9660'.

Technical References

APPLE COMPUTER COMPANY (1989). Apple CD-Rom SC developers guide, revised edition. AAPDA Document M211. (AAPDA; Sydney).

ATSATT, B. and BECHTEL, B. (1989). Apple extensions to ISO 9660. *In* Bechtel (1989) *op. cit.*

BECHTEL, B. (1989). CD-Rom and the Macintosh computer. Apple Advanced Technology Group report. Apple Technical Report No. 16. (Apple Computer Company; Cupertino).

DISCTRONICS (1989). Disctronics compact disc handbook. Release 4, Australia and Pacific. Melbourne.

ISO (1988). Volume and file structure of CD-Rom for information interchange. First Edition: 1988-04-15.

APPENDIX C. A clarification checklist

1. What Standard do you want the CD-Rom created to?
   [High Sierra? Mac HFS, Apple ProDos? ISO 9660? Note the spoecial requirements for full support on different machines and under different operating systems]

2. Do you wish to create an XA disc?
   [if so, this report merely warns that you should get further specialised advice to decide if this is indeed what you require]

3. Do you want to simply archive data or do you want an embedded program on the CD-Rom to handle fast access indexing?
   [If this is the case, choose with care: some products require special versions for different computer types, and you may limit your market uncessarily- and unintentionally]

4. Do you want potential users without CD-Rom drives to be able to get something from your Rom?
   [If so, consider adding an audio track accessible to ordinary CD audio players as track 1 or more

ARRB
Information
Resource
Rom

GREEN
disc, with
silver
rings