

Joliet Specification

CD-ROM Recording Spec ISO 9660:1988

Extensions for Unicode Version 1; May 22, 1995

Copyright 1995, Microsoft Corporation All Rights Reserved
Contact Microsoft Developer Relations Group
MAC@avca.com

CONTENTS

- [Preface](#)
 - [Purpose and Scope](#)
 - [Overview](#)
 - [Terminology and Notation](#)
- [Joliet Recording Specification](#)
 - [Change Summary](#)
- [Identifying an ISO 9660 SVD as Unicode \(UCS-2\)](#)
 - [SVD Escape Sequences Field](#)
 - [SVD Volume Flags Field](#)
 - [Resolution of ISO 9660 Ambiguities for Wide Characters](#)
 - [Wide Character Byte Ordering](#)
 - [Allowed Character Set](#)
 - [Special Directory Identifiers](#)
 - [Separator Characters](#)
 - [Sort Ordering](#)
 - [Relaxation of ISO 9660 Restrictions on UCS-2 Volumes](#)
- [Extensions to Joliet](#)
 - [Joliet for Multisession Media](#)
 - [CD-XA Extensions to Joliet](#)
 - [Other Extensions to Joliet](#)
- [Bibliography 14](#)

Preface

Purpose and Scope

While the CD-ROM media provides for cost-effective software distribution, the existing ISO

9660 file system contains a number of restrictions which interfere with simple and efficient distribution of files on a CD-ROM.

The read-only nature of CD-ROM media has led content authors to continue to use traditional magnetic media as their main avenue for creating applications. Each of the existing file systems for magnetic media contain various features which can not be represented on CD-ROM media using an unenhanced version of ISO 9660.

As content authors attempt to transfer their applications to the CD-ROM, they are likely to find that some of their work cannot be distributed on the CD-ROM media due to restrictions in the ISO 9660 format. This frustrates some content authors.

Because the CD-ROM media is mainly a distribution media, rather than a creative (read/write) media, it is necessary for the CD-ROM file system to support a superset of the creative media features. This fundamental flaw in the design of ISO 9660 has prompted several operating systems vendors to extend ISO 9660 in several ways. Some examples are Rock Ridge Interchange Protocol and Apple's use of the System Use Area to store finder flags.

Some of the ISO 9660 problems which are addressed by this specification include:

- Character Set limitations.
- File Name Length limitations
- Directory Tree Depth limitations
- Directory Name Format limitations
- Wide Character (16-bit character) ambiguities

The general design approach used in the Joliet specification is to relax restrictions and resolve ambiguities in the ISO 9660:1988 specification so the practical goals can be met.

Overview

The Joliet specification utilizes the supplementary volume descriptor (SVD) feature of ISO 9660 to specify a set of files recorded within the Unicode character set.

The ISO 10646 character set specification may be identified by an ISO 2022 escape sequence. By recording this escape sequence in an ISO 9660 SVD, this technique for identifying the Unicode SVD is compliant with the ISO 9660 specification. It also retains interchange by not disrupting the files referenced through the primary volume descriptor (PVD).

All that remains is to resolve minor technical ambiguities within ISO 9660 which arise as the result of the use of wide characters.

Because the use of this particular escape sequence in an ISO 9660 SVD is unprecedented up to this time, several of the restrictions which are imposed by ISO 9660 may be relaxed without significantly disrupting information interchange between existing systems from a practical

standpoint.

This design approach has several benefits. For instance, the use of the existing ISO 9660 standard allows for straightforward integration with existing extensions to ISO 9660. The designs for the System Use Sharing Protocol, Rock Ridge extensions for POSIX semantics, CD-XA System Use Area Semantics, Apple's Finder Flags and Resource Forks, all port in a straightforward manner to the Joliet specification.

Also, the use of a new SVD eliminates the danger of breaking software compatibility with existing ISO 9660 systems. Existing software will simply ignore the Unicode SVD, and will simply use the PVD instead. This compatibility "safety-valve" makes the goal of relaxing the file system's restrictions easier.

This document describes how a CD-ROM may be constructed so that names on the volume can be recorded in Unicode while remaining in compliance with ISO 9660. The particular ISO 10646 character sets used here are UCS-2 Level 1, UCS-2 Level 2, and UCS-2 Level 3.

The basic strategy of CD-ROM volume recognition is the Volume Recognition Sequence, which is a sequence of volume descriptors, recorded one per sector, starting at Sector 16 in the first track of the last session on the disc. A receiving system reads these sectors and chooses a particular volume descriptor from the sequence. This volume descriptor acts as a kind of anchor upon which the remainder of the volume is constructed.

Terminology and Notation

Joliet is based on the ISO 9660:1988 standard. Unless defined in this document, the terminology used shall be as defined in ISO 9660:1988.

The following notation is used in this document.

- Decimal and Hexadecimal Notation
 - Numbers in decimal notation are represented by decimal digits, namely 0 to 9.
 - Numbers in hexadecimal notation are represented by hexadecimal digits, namely 0 to 9 and A to F, shown in parentheses. For instance, the hexadecimal number D0 shall be written as (D0).
- A literal sequence of ASCII characters will be represented by those characters within single quotes. For instance, 'ABC' means the byte sequence (41)(42)(43).
- References to characters in the ISO 2022 escape sequence will be given in comma-separated decimal nibble/nibble format, in hexadecimal format, and as ASCII characters, with equal signs between each format, all enclosed within parenthesis. For instance, the 3-byte ISO 2022 escape sequence for Shift-JIS is (2/4, 2/11, 3/10 = (24)(2B)(3A)= '\$+:').

[return to the table of contents](#)

Joliet Recording Specification

Change Summary

The Joliet specification resolves the following ISO 9660 ambiguities for UCS-2 volumes:

- Use a SVD with a UCS-2 (UNICODE) Escape Sequence.
- The UCS-2 escape sequences used are: (25)(2F)(40), (25)(2F)(43), or (25)(2F)(45).
- The default setting of bit 0 of the SVD "Volume Flags Field" is ZERO.
- The Unicode Wide characters shall be recorded in "Big Endian" (Motorola) format.
- Special Directory Identifiers are recorded as single byte names containing (00) or (01).
- SEPARATOR 1 and SEPARATOR 2 are encoded using an equivalent 16-bit code point.
- Sort ordering is unchanged, except that all justification pad bytes are to be set to (00).

The Joliet specification recommends that several ISO 9660 restrictions be lifted on UCS-2 volumes. The Joliet specification allows for the following interchange rules:

- The File or Directory Identifiers may be up to 128 bytes (64 unicode characters) in length.
- Directory Identifiers may contain file name extensions.
- The Directory Hierarchy may be recorded deeper than 8 levels.
- The volume recognition sequence supports multisession. This is compatible with the CD-Bridge specification.

The Joliet specification may be extended through the use of the following specifications:

- Mode 2 Form 2 extents and CD-DA extents, ("System Description CD-ROM XA")
- System Use Sharing Protocol (not explicitly specified here)
- RockRidge Interchange Protocol (not explicitly specified here)
- Other future CD-ROM file system formats

[return to the table of contents](#)

Identifying an ISO 9660SVD as Unicode (UCS-2)

SVD Escape Sequences Field

The Escape Sequences field of an ISO 9660 Supplementary Volume Descriptor (ISO 9660 section 8.5.6) shall identify the character set used to interpret descriptor fields related to the Directory Hierarchy identified by the Volume Descriptor.

If the Escape Sequences field of an ISO 9660 SVD identifies any of the following UCS-2 escape sequences, then the descriptor fields related to the Directory Hierarchy identified by that Volume Descriptor shall be interpreted according to the identified UCS-2 character set.

Table 1 - ISO 2022 UCS-2 Escape Sequences

ISO 2022 Escape Sequence as recorded in the ISO 9660 SVD

Standard	Level	Decimal	Hex Bytes	ASCII
UCS-2	Level 1	2/5, 2/15, 4/0	(25)(2F)(40)	'%\@'
UCS-2	Level 2	2/5, 2/15, 4/3	(25)(2F)(43)	'%\C'
UCS-2	Level 3	2/5, 2/15, 4/5	(25)(2F)(45)	'%\E'

A "Unicode Volume" refers to the Volume Descriptor and Directory Hierarchy identified by a Supplementary Volume Descriptor containing an Escape Sequences field which identifies any of the above UCS-2 character sets.

SVD Volume Flags Field

The UCS-2 Level 1, UCS Level 2, and UCS-2 Level 3 escape sequences are considered to be registered according ISO 2735 for purposes of setting bit 0 of the Volume Flags field of the SVD.

The nominal value of Bit 0 of the Volume Flags field for a Unicode SVD shall be ZERO.

Resolution of ISO 9660 Ambiguities for Wide Characters

This specification resolves ISO 9660 ambiguities with respect to wide (16-bit) character sets, such as the UCS-2 character set.

Wide Character Byte Ordering

All UCS-2 characters shall be recorded according to ISO 9660:1988 section 7.2.2, 16-bit numerical value, most significant byte first ("Big Endian").

Allowed Character Set

All UCS-2 code points shall be allowed except for the following UCS-2 code points:

- All code points between (00)(00) and (00)(1F), inclusive. (Control Characters)
- (00)(2A) '*' (Asterisk)
- (00)(2F) '/' (Forward Slash)
- (00)(3A) ':' (Colon)
- (00)(3B) ';' (Semicolon)
- (00)(3F) '?' (Question Mark)
- (00)(5C) '\' (Backslash)

[return to the table of contents](#)

Special Directory Identifiers

Section 7.6 of ISO 9660 describes the recording of reserved directory identifiers for the root, current, and parent directory identifiers as single (00) or single (01) bytes.

In a wide character set, it is not possible to represent a character in a single byte. The following portions of the ISO 9660:1988 specification referring to reserved directory identifiers are ambiguous.

The ISO 9660:1988 sections in question are as follows:

- 6.8.2.2 Identification of directories
- 7.6.2 Reserved Directory Identifiers
- 9.1.11 File Identifier
- 9.4.5 Directory Identifier

These special case directory identifiers are not intended to represent characters in a graphic character set. These characters are placeholders, not characters. Therefore, these definitions remain unchanged on a volume recorded in Unicode.

Simply put, Special Directory Identifiers shall remain as 8-bit values, even on a UCS-2 volume, where other characters have been expanded to 16-bits.

Root Directory

The Directory Identifier of a Directory Record describing the Root Directory shall consist of a single (00) byte.

Current Directory

The Directory Identifier of the first Directory Record of each directory shall consist of a single (00) byte.

Parent Directory

The Directory Identifier of the second Directory Record of each directory shall consist of a single (01) byte.

Separator Characters

The separator characters SEPARATOR 1 and SEPARATOR 2 are specified as 8-bit characters, which can not be represented in a wide character set, so the ISO 9660:1988 specification sections referring to SEPARATOR 1 and SEPARATOR 2 are ambiguous.

The ISO 9660:1988 sections in question are as follows:

- 7.4.3 Separators
- 7.5.1 File Identifier format
- 7.5.2 File Identifier length
- 8.4.24 Abstract File Identifier
- 8.4.25 Bibliographic File Identifier
- 8.5.17 Copyright File Identifier
- 8.5.19 Bibliographic File Identifier

- 9.1.11 File Identifier

The values SEPARATOR 1 and SEPARATOR 2 shall be represented differently depending on the d1 character set.

In the case of an SVD identifying a UCS-2 character set, the values of SEPARATOR 1 and SEPARATOR 2 shall be recorded as a UCS-2 character with an equivalent code point value.

Otherwise, the definitions of SEPARATOR 1 and SEPARATOR 2 shall be recorded according to section 7.4.3 of ISO 9660:1988.

Simply put, SEPARATOR 1 and SEPARATOR 2 shall be expanded to 16-bits.

Table 2 - Separator Representations

ISO 9660:1988 Volume Unicode Volume

Separator	Bit Combination	UCS-2 Codepoint
SEPARATOR 1	(2E)	(00)(2E)
SEPARATOR 2	(3B)	(00)(3B)

[return to the table of contents](#)

Sort Ordering

ISO 9660 specifies the order of path table records within a path table, and specifies the order of directory records within a directory. These sorting algorithms assume an 8-bit character set is used. These sorting algorithms are ambiguous when used with wide characters.

The ISO 9660:1988 sections in question are as follows:

- 6.9.1 Order of Path Table Records
- 9.3 Order of Directory Records

The only change required is to redefine the value of the sort justification pad byte to zero (00).

Simply put, comparing the byte contents in all positions remains a suitable sorting algorithm for the descriptor fields recorded in a UCS-2 SVD Directory Hierarchy. This is one of the primary reasons for selecting the Big Endian format to represent all UCS-2 characters.

Natural Language Sorting

On a Unicode volume, the 16-bit UCS-2 code points are used to determine the Order of Path Table Records and the Order of Directory Records.

No attempt will be made to provide natural language sorting on the media. Natural language

sorting may optionally be provided by a display application as desired.

Justification Pad Bytes

The sort ordering algorithms as specified in ISO 9660:1988 sections 6.9.1 and 9.3 are acceptable except for the value of the justification "pad byte".

The value of the justification "pad byte" as specified in ISO 9660:1988 section 6.9.1 shall be (00). This is changed from a value of (20) as specified in that same section.

The value of the justification "pad byte" as specified in ISO 9660:1988 section 9.3 subsections (a) and (b) shall be (00). This is changed from a value of (20) as specified in those same sections.

The value of the justification "pad byte" as specified in ISO 9660:1988 section 9.3 subsections (c) shall be (00). This is changed from a value of (30) as specified in that same section.

Simply put, set all the justification "pad bytes" to zero to simplify sorting.

Mandatory Sort Ordering.

Correct sort ordering is mandatory on UCS-2 volumes.

Descriptor Fields affected by the UCS-2 Escape Sequence

If a UCS-2 escape sequence is detected in a supplementary volume descriptor, the following descriptor fields referenced from that supplementary volume descriptor shall contain UCS-2 characters.

- ISO 9660:1988 Section 8.5.4 System Identifier
- ISO 9660:1988 Section 8.5.5 Volume Identifier
- ISO 9660:1988 Section 8.5.13 Volume Set Identifier
- ISO 9660:1988 Section 8.5.14 Publisher Identifier
- ISO 9660:1988 Section 8.5.15 Data Preparer Identifier
- ISO 9660:1988 Section 8.5.16 Application Identifier
- ISO 9660:1988 Section 8.5.17 Copyright File Identifier
- ISO 9660:1988 Section 8.5.18 Abstract File Identifier (missing section)
- ISO 9660:1988 Section 8.5.19 Bibliographic File Identifier
- ISO 9660:1988 Section 9.1.11 File Identifier
- ISO 9660:1988 Section 9.4.5 Directory Identifier
- ISO 9660:1988 Section 9.5.11 System Identifier (of Extended Attribute Record)

[return to the table of contents](#)

Relaxation of ISO 9660 Restrictions on UCS-2 Volumes

Several ISO 9660 restrictions will be relaxed to achieve a more useful recording specification.

Joliet receiving systems shall be capable of receiving media recorded with restrictions which have been relaxed relative to ISO 9660.

Maximum File Identifier Length Increased

Joliet receiving systems shall receive directory hierarchies recorded with file identifiers longer than those allowed by ISO 9660 receiving systems.

ISO 9660 (Section 7.5.1) states that the sum of the following shall not exceed 30:

- If there is a file name, the length of the file name.
- If there is a file name extension, the length of the file name extension.

On Joliet compliant media, however, the sum as calculated above shall not exceed 128, to allow for longer file identifiers.

The above lengths shall be expressed as a number of bytes.

Maximum Directory Identifier Length Increased

Joliet receiving systems shall receive directory hierarchies recorded with file names longer than those allowed by ISO 9660 receiving systems.

ISO 9660 (Section 7.6.3) states that the length of a directory identifier shall not exceed 31.

On Joliet compliant media, however, the length of a directory identifier shall not exceed 128, to allow for longer directory identifiers.

The above lengths shall be expressed as a number of bytes.

Directory Names May Have File Name Extensions

ISO 9660 does not allow directory identifiers to contain file name extensions.

On Joliet compliant media, however, directory identifiers may contain file name extensions.

The Joliet directory identifier format shall be calculated according to ISO 9660 section 7.5.1 "File Identifier format", with the exception that the length of a directory identifier may exceed 31, but shall not exceed 128.

In addition, the Joliet directory identifier format shall comply with ISO 9660 section 7.6.2 "Reserved Directory Identifiers".

The directory identifier length shall be calculated according to ISO 9660 section 7.5.2 "File Identifier length".

The above lengths shall be expressed as a number of bytes.

Maximum Directory Hierarchy Depth May Exceed 8 Levels

ISO 9660 (Section 6.8.2.1) specifies restrictions regarding the Depth of Directory Hierarchy. This section of ISO 9660 specifies that this number of levels in the hierarchy shall not exceed eight.

On Joliet compliant media, however, the number of levels in the hierarchy may exceed eight.

Joliet compliant media shall comply with the remainder of ISO 9660 section 6.8.2.1, so that for each file recorded, the sum of the following shall not exceed 240:

- the length of the file identifier;
- the length of the directory identifiers of all relevant directories;
- the number of relevant directories.

The above lengths shall be expressed as a number of bytes.

[return to the table of contents](#)

Extensions to Joliet

Joliet for Multisession Media

Multisession Recordings are Received

When provided with CD-ROM reader hardware with multisession capability, Joliet receiving systems shall receive media recorded using the multisession recording technique.

The details of this technique are provided below

Logical Sector Addressing on Multisession Recordings

Each sector on the media is assigned a unique Logical Sector Address.

Logical Sector Addresses zero and above increase linearly across the surface of the disc, regardless of session boundaries.

Logical Sector Address zero references the sector with Minute:Second:Frame address 00:02:00 in the first session. All other Logical Sector Addresses are relative to Minute:Second:Frame address 00:02:00 in the first session.

The conversion between Logical Sector Addresses and Minute:Second:Frame addresses is
 Logical Sector Address = (((Minute*60)+Seconds)*75) - 150.

Simply put, the Logical Sector Address on a multisession disc describes a flat address space.

Multisession Addressability

The data area for a volume may span multiple sessions.

For example, if a disc is recorded with 3 sessions, the directory hierarchy described by a volume descriptor in session 3 may reference logical sectors recorded in session 1, 2, or 3.

Multisession Volume Recognition Sequence

The Volume Recognition Sequence shall begin at the 16th logical sector of the first track of the last session on the disc.

This volume recognition sequence supersedes all other volume recognition sequences on the disc. The interpretation of the Volume Recognition Sequence is otherwise unchanged.

For example, consider a disc that contains 3 sessions, where session 1 starts at 00:00:00, session 2 starts at 10:00:00, and session 3 starts at 20:00:00. The Volume Recognition Sequence for this disc would start at Minute:Second:Frame address 20:00:16.

This technique is compatible with the CD-Bridge multisession technique.

Track Modes and Sector Forms

The data area for a Joliet volume on a CD-ROM shall be comprised of either Mode 1 or Mode 2 Form 1 sectors. CD-ROM media utilizing the multisession recording techniques outlined above may not contain any Mode 1 sectors anywhere on the media. Mode 1 sectors are allowed only on single-session media.

Mode 2 Form 2 sectors and CD-Digital Audio tracks may be recorded on the same media as a Joliet volume. In this case, the CD-XA extensions to Joliet may be utilized to identify Mode 2 Form 2 extents and CD-Digital Audio extents.

CD-Digital Audio tracks may not be recorded in sessions 2 and higher. If any CD-Digital Audio tracks are recorded, all the CD-Digital Audio tracks shall be recorded in the first session.

CD-XA Extensions to Joliet

CD-ROM discs utilizing the Joliet extensions to ISO 9660 and which also identify mode 2 form 2 extents or CD-Digital Audio extents shall be marked with a CD-ROM XA Label as specified in "System Description CD-XA" section 2.1.

The CD-ROM XA Label shall be located at offset 1024 (byte position 1025) in the Joliet Supplementary Volume Descriptor. The identifying signature 'CD-XA001' shall be recorded starting at offset 1024 in the Joliet Supplementary Volume Descriptor. This identifying signature is equivalent to the hex bytes (43)(44)(2D)(58)(41)(30)(30)(31).

Mode 2 form 2 extents shall be identified using recording rules outlined in "System Description

CD-XA", section 2.7. In this case, bit 12 of the Attributes field of the "XA System Use Information" shall be set to one to identify that the file contains mode 2 form 2 sectors. See below for additional information regarding Data Length.

CD-Digital Audio extents shall be identified using recording rules outlined in "System Description CD-XA", section 2.7. In this case, bit 14 of the Attributes field of the "XA System Use Information" shall be set to one to identify that the file is comprised of an extent of CD-Digital Audio. See below for additional information regarding Data Length.

If a file is marked such that either bit 12 is set to one or bit 14 is set to one in the Attributes field of the "XA System Use Information", then the Data Length field of the Directory Record shall be set to 2048 times the number of sectors contained in the extent.

See ISO 9660:1988 section 9.1.4.

Other Extensions to Joliet

The Joliet Extensions to ISO 9660 are designed to coexist with other extensions such as the "System Use Sharing Protocol" and "RockRidge Interchange Protocol". However, these protocols are not an integral part of the Joliet specification.

The method used to integrate these other protocols into Joliet is not defined here.

[return to the table of contents](#)

Bibliography

ISO 2022 - Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques, International Organization for Standardization,

ISO 9660 - Information processing - Volume and file structure of CD-ROM for information interchange, International Organization for Standardization, 1988-04-15

ISO 10149 : 1989 (E) - Information technology - Data interchange on read-only 120mm optical data discs (CD-ROM) "YellowBook", International Organization for Standardization, 1989-09-01

ISO 10646 - Information technology - Universal Multiple-Octet Coded Character Sets (UCS), International Organization for Standardization,

The Unicode Standard - Worldwide Character Encoding Version 1.0, The Unicode Consortium, Addison-Wesley Publishing Company, Inc, 1990-1991 Unicode, Inc., Volume 1

Orangebook, N. V. Philips and Sony Corporation, November 1990

System Description CD-XA, N. V. Philips and Sony Corporation, March 1991

System Use Sharing Protocol

RockRidge Interchange Protocol

Copyright © 1995 Microsoft Corporation unless otherwise specified. All Rights Reserved.