

Embedded font checks for PDF/A validation

Version 0.1, January 10, 2015

1. Normative references

PDF/A specifications: ISO 19005-1:2005, ISO 19005-2:2011, ISO 19005-3:2012.

PDF Specifications: PDF 1.4 Specification (Adobe), ISO 32000-1:2008.

Font specifications:

- [1] Apple Computer, Inc., TrueType Reference Manual. Available on Apple's Web site at <http://developer.apple.com/fonts/TTRefMan/>
- [2] Microsoft Corporation, TrueType 1.0 Font Files Technical Specification. Available at <http://www.microsoft.com/typography/tt/tt.htm>
- [3] Microsoft Corporation, OpenType specification, version 1.6. Available at <http://www.microsoft.com/typography/otspec/>
- [4] Open Font Format, ISO/IEC 14496-22:2009 (Second Edition).
- [5] Adobe Type 1 Font Format, Adobe Systems Incorporated, ISBN 0-201-57044-0, 1990.
- [6] Technical Note #5015, Type 1 Font Format Supplement, 15 January 1994, Adobe Systems Incorporated.
- [7] Technical Note #5088, Font Naming Issues, 12 April 1993, Adobe Systems Incorporated.
- [8] Technical Note #5092, CID-Keyed Font Technology Overview, Adobe Developer Support, 12 September 1994, Adobe Systems Incorporated.
- [9] Technical Note #5176, The Compact Font Format Specification, Version 1.0, 18 March 1998, Adobe Systems Incorporated.
- [10] Technical Note #5177, The Type 2 Charstring Format, 5 May 1998, Adobe Systems Incorporated.
- [11] Technical Note #5641, Enabling PDF Font Embedding for CID-Keyed Fonts, 7 July 1998, Adobe Systems Incorporated.
- [12] PostScript Language Reference, Third Edition, Adobe Systems Incorporated, ISBN 0-201-37922-8, 1999.

2. Terminology

PDF Font Dictionary: either a simple font dictionary as specified by ISO 32000-1:2008, 9.6 or a CIDFont dictionary as specified in ISO 32000-1:2008, 9.7.4.

PDF Font Type: the value of the key **Subtype** in the PDF Font Dictionary.

PDF Font Descriptor Dictionary: a font descriptor dictionary referred by **FontDescriptor** key in the PDF Font Dictionary and specified in ISO 32000-1:2008, 9.8.

PDF Font File Stream: a PDF stream containing the embedded font program (file) referred by one of the keys **FontFile**, **FontFile2**, **FontFile3** in the PDF Font Descriptor Dictionary and specified in ISO 32000-1:2008, 9.9.

PostScript Type1 Font: Type1 font file in PostScript format as specified by [5].

CharStrings Dictionary: A dictionary associating character names (the keys in the dictionary) with glyph descriptions in PostScript Type1 Font file. See [5], Chapter 6.

CFF Font: Type1 or CIDFontType0 font file in compact font format as specified by [9, 10].

CharStrings INDEX: an array of all glyph descriptions in CFF Font. See [9], Section 14.

GID: an index used to identify glyph description either in the Charstrings INDEX of the CFF Font or in “glyf” table of the TrueType or OpenType Font.

CFF Charset: A structure in the CFF Font providing the mapping from the glyph names to GIDs. See [9], Section 13.

TrueType Font: TrueType font file as specified by technically equivalent documents [1, 2].

OpenType Font: OpenType font file as specified by technically equivalent documents [3, 4].

TrueType/OpenType table: a named byte range of the TrueType or OpenType font file as defined in [1-4].

3. Formats of embedded font files

PDF specification supports the following formats of embedded font files:

PDF Font Type	Key in the PDF Font Descriptor Dictionary	Value of Subtype key in the PDF Font File Stream	Font file format	Normative reference
Type1, MMTYPE1	FontFile	-	PostScript Type1	[5, 12]
Type1, MMTYPE1	FontFile3	Type1C	CFF	[9]
Type1	FontFile3	OpenType	OpenType with “CFF” table	[3,4,9]
TrueType	FontFile2	-	TrueType	[1,2]
TrueType	FontFile3	OpenType	OpenType with “glyf” table	[3,4]
CIDFontType0	FontFile3	CIDFontType0C	CFF	[9]
CIDFontType0	FontFile3	OpenType	OpenType with “CFF” table	[3,4,9]
CIDFontType2	FontFile2	-	TrueType	[1,2]
CIDFontType2	FontFile3	OpenType	OpenType with “glyf” table	[3,4]

4. Embedded font file requirements

4.1. PostScript Type1 Fonts

4.1.1. The values of keys **Length1**, **Length2**, **Length3** of the PDF Font File Stream shall be correct.

4.1.2. The general font file organization shall comply to [5], Chapter 2.

4.1.3. The glyph with name “.notdef” shall be present in the CharStrings dictionary.

4.1.4. The font file dictionary shall contain a valid Encoding array as specified by [5], 2.2 and [12], 5.3.

4.1.5. Names of all glyphs referenced for rendering shall be present in the CharStrings dictionary. A glyph name is referenced for rendering if it is mapped from the character referenced for rendering via the Encoding mechanism for Type1 fonts as specified by ISO-32000:1, 9.6.6.2.

4.1.6. If the **CharSet** key is present in the PDF Font Descriptor Dictionary, the names of all glyphs specified in its value shall be present in the font CharStrings dictionary, regardless of whether this glyph is referenced for rendering or not.

4.1.7. Charstrings for all glyphs in 4.1.2, 4.1.4, 4.1.5 shall comply with the charstring encoding specification in [5], Chapter 6.

4.1.8. Glyph widths referenced for rendering shall be consistent with the width information in PDF Font Dictionary. Glyph widths in the PostScript Type1 file are determined by the Metrics dictionary of the font file (see [5], 2.2; [12], 5.9.2) or, if it is not present, by “hsbw” or “sbw” operator in the glyph charstring (see [5], 6.4).

Q1. Do we impose this condition only for glyphs referenced for rendering or, in addition, for all glyphs referenced by character codes from [FirstChar, LastChar] range?

4.2. Compact Font File (CFF) in case of Type1 and MMTYPE1 PDF Font Types

4.2.1. The general CFF Font file structure shall comply to [9], Section 2 and shall consist only of a single font. In particular, it shall contain a valid Header, Name INDEX, Top DICT INDEX, String INDEX, Global Subr INDEX, Encoding, Charset, CharStrings INDEX, Font DICT INDEX, Private DICT.

4.2.2. All GIDs referenced for rendering from CIDs via the algorithm defined in ISO 32000-1:2008, 9.7.4.2 shall be present in the CharStrings INDEX.

4.2.3. If the **CharSet** key is present in the PDF Font Descriptor Dictionary, the names of all glyphs specified in its value shall be present in the Charset structure, regardless of whether this glyph is referenced for rendering or not.

4.2.4. GIDs for all glyphs in 4.2.2, 4.2.3 identified via Charset structure, shall point to valid charstrings in the CharStrings INDEX as specified by [9], Section 14; [10].

4.2.5. Glyph widths referenced for rendering shall be consistent with the width information in PDF Font Dictionary. Glyph widths in the CFF file are determined by “hsbw” or “sbw” operator in the glyph charstring of Type1, or as a first number of the Type2 charstring serving as a difference to *nominalWidthX*, or, if omitted, as a *defaultWidthX* (see [10], 3.1). The *nominalWidthX* and *defaultWidthX* are defined in the Private DICT of the CFF font.

4.3. Compact Font File (CFF) in case of CIDFontType0 PDF Font Type

4.2.1. The general CFF Font file structure shall comply to [9], Section 2 and shall consist only of a single font. In particular, it shall contain a valid Header, Name INDEX, Top DICT INDEX, String INDEX, Global Subr INDEX, CharStrings INDEX, Font DICT INDEX, Private DICT and, optionally, the FDSelect structure. Both the Encoding and the Charset structures are optional and are not used for locating glyph charstrings.

4.2.2. All GIDs mapped from CIDs used for rendering via an algorithm defined in 9.7.4.2 shall be present in the font file and correctly encoded.

4.2.3. Glyph widths referenced for rendering shall be consistent with the width information in PDF Font Dictionary. Glyph widths in the CFF file are determined by “hsbw” or “sbw” operator in the glyph charstring of Type1, or as a first number of the Type2 charstring serving as a difference to

nominalWidthX, or, if omitted, as a defaultWidthX (see [10], 3.1). The nominalWidthX and defaultWidthX are defined in the Private DICT of the CFF font.

4.4. TrueType Font File in case of TrueType PDF Font Type

4.4.1. The font shall contain the following minimal set of tables: “cmap”, “glyf”, “head”, “hhea”, “hmtx”, “loca”, “maxp”. The “cvt”, “fpgm”, and “prep” tables must also be included if they are required by the font instructions. All these tables shall comply to the data format requirements of [1,2].

Q2. The tables “name” and “post” are required by TrueType specification, but are not required by ISO 32000-1:2008. What takes precedence here?

4.4.2. If the PDF Font Descriptor **Flags** key identifies the PDF Font as non-symbolic (ISO 32000-1:2008, 9.8.2) and the PDF Font Encoding defines Differences array, then the “cmap” table shall contain at least Microsoft Unicode (3,1 – Platform ID=3, Encoding ID=1) encoding.

4.4.3. In case of PDF/A-1 standard, if the PDF Font Descriptor Flags key identifies the PDF Font as symbolic (ISO 32000-1:2008, 9.8.2), then the “cmap” table shall contain exactly one encoding.

4.4.4. In case of either PDF/A-2 or PDF/A-3 standard, if the PDF Font Descriptor Flags key identifies the PDF Font as symbolic (ISO 32000-1:2008, 9.8.2), then the “cmap” table shall either contain exactly one encoding or at least Microsoft Symbol (3,0 – Platform ID=3, Encoding ID=0) encoding. If Microsoft Symbol encoding is present, the range of character codes shall be one of these: 0x0000 - 0x00FF, 0xF000 - 0xFFFF, 0xF100 - 0xF1FF, or 0xF200 - 0xF2FF.

4.4.5. All GIDs of glyphs used for rendering, as determined by the algorithm described in ISO 32000-1, 9.6.6.4, shall be present in the “glyf” table and their instructions shall comply with [1,2].

4.4.6. All GIDs of glyphs used for rendering, as determined by the algorithm described in ISO 32000-1, 9.6.6.4, shall be present in the “hmtx” table and their widths shall be consistent with widths information of the PDF Font Dictionary.

Q3. The algorithm described in ISO 32000-1, 9.6.6.4 “Encodings for TrueType Fonts” is not consistent with PDF/A requirements in case of symbolic fonts. It looks like this algorithm should be adjusted as follows:

- If the font contains a (3, 0) subtable, the range of character codes shall be one of these: 0x0000 - 0x00FF, 0xF000 - 0xFFFF, 0xF100 - 0xF1FF, or 0xF200 - 0xF2FF. Depending on the range of codes, each byte from the string shall be prepended with the high byte of the range, to form a two-byte character, which shall be used to select the associated glyph description from the subtable.
- Otherwise, if the “cmap” table contains a (1,0) subtable, single bytes from the string shall be used to look up the associated glyph descriptions from this subtable
- If the font contains exactly one encoding different from (3,0) and (1,0) subtypes, single bytes from the string shall be used to look up the associated glyph descriptions from this subtable.

4.5. TrueType Font File in case of CIDFontType2 Font Type

4.5.1. The font shall contain the following minimal set of tables: “glyf”, “head”, “hhea”, “hmtx”, “loca”, “maxp”. The “cvt”, “fpgm”, and “prep” tables must also be included if they are required by the font instructions. The tables “vhea” and “vmtx” shall also be included if the PDF Font is used for vertical writing. All these tables shall comply to the data format requirements of [1,2].

4.4.5. All GIDs of glyphs used for rendering, as determined by the algorithm described in ISO 32000-1, 9.7.4.2, shall be present in the “glyf” table and their instructions shall comply with [1,2].

4.4.6. All GIDs of glyphs used for rendering, as determined by the algorithm described in ISO 32000-1, 9.7.4.2, shall be present in the “hmtx” table and their widths shall be consistent with widths information of the PDF Font Dictionary (keys **W** and **DW**). If the PDF Font is used for vertical writing, the same condition applies to “vmtx” table and metrics information of the PDF Font Dictionary (keys **W2** and **DW2**).

Q4. Do we really need to require the consistency of metrics used for vertical writing?

4.6. OpenType fonts

4.6.1. An OpenType font file shall not contain both “glyf” and “CFF ” tables.

4.6.2. If an OpenType font file contains “glyf” table shall comply either with the requirements of Section 4.4 in case of TrueType PDF Fonts or with the requirements of Section 4.5 in case of CIDFontType2 PDF Fonts.

4.6.3. If an OpenType font file contains “CFF ” table, its data shall comply either with the requirements of Section 4.2 in case of Type1 or MMTyep1 PDF Fonts or with the requirements of Section 4.3 in case of CIDFontType0 PDF Fonts.

Q5. ISO 32000-1:2008 requires the presence of “cmap” table along with “CFF ” table. It doesn’t seem to be required at all?