PDF association

Specification

Brotli compression in PDF 2.0

EXTN-BROTLI-1 v1.3

PDF TWG

E-mail: copyright@pdfa.org
Web: https://pdfa.org

Table of Contents

Introduction

PDF is a digital format for representing page-based documents. PDF files are made up of a set of objects, and indexes for those objects. These objects can represent a range of different things, including streams of graphical operators, image and font data, and other such resources. PDF supports various different compression methods that can be applied to these objects to save storage space.

Some methods (specifically those aimed at image compression) are lossy, in that quality is traded off against size. Other more general purpose methods (which can operate on any given resource, not just images) are lossless; if a stream of data is compressed by such a method, the result of decompressing it will be byte for byte identical to the initial data.

When PDF first appeared, over 3 decades ago, the primary lossless compression method provided was LZW. This was followed shortly thereafter (in PDF 1.2) by the addition of Flate.

Over the intervening time new and better algorithms have become available, promising improved compression ratios for comparable processing requirements. This extension adds support for such a modern lossless compression algorithm, Brotli.

Brotli is a new, state of the art standard developed for use on the web. A development of the same LZ77 algorithm that underlies the Flate algorithm, Brotli typically offers faster speeds for similar compression ratios. Higher compression settings can offer significantly better compression, albeit at a cost of noticeably longer compression times.

Brotli is a well-standardised, patent-free specification with several independent implementations.

Every new compression algorithm imported into PDF imposes a cost on every PDF processor; it is therefore preferable to pick a single candidate rather than to force resource-constrained implementations to have to add multiple new decoders.

This document follows existing conventions for PDF specifications, as specified in ISO 32000, Clauses 4 to 6. The PDF Association provides sponsored editions of many ISO standards for no-cost.

1 Scope

This document specifies an extension to the PDF 2.0 specification (ISO 32000) to allow the use of Brotli for compression of stream data. This takes the form of a new filter that can be applied to stream objects.

This document does not specify the following:

- specific processes for converting paper or electronic documents to the PDF file format;
- specific technical design, user interface implementation, or operational details of rendering;
- specific physical methods of storing these documents such as media and storage conditions;
- methods for validating the conformance of PDF files or PDF processors;
- required computer hardware or operating system.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 32000 (all parts), International Organization for Standardization (committee). *Document management—Portable document format*. First edition. Geneva: International Organization for Standardization. https://www.iso.org/standard/51502.html.

IETF RFC 7932, ALAKUIJALA, J. and Z. SZABADKA. *Brotli Compressed Data Format*. 2016. RFC Publisher. https://www.rfc-editor.org/info/rfc7932.

IETF RFC 9841, ALAKUIJALA, J., T. DUONG, E. KLIUCHNIKOV and Z. SZABADKA. *Shared Brotli Compressed Data Format*. 2025. RFC Publisher. https://www.rfc-editor.org/info/rfc9841.

3 Definitions

For the purposes of this document, the terms and definitions given in ISO 32000 apply.

The PDF Association maintains terminology databases for use in describing the Portable Document Format at the following addresses:

- Glossary of general PDF terms: available at https://pdfa.org/glossary-of-pdf-terms/
- Glossary of PDF accessibility terminology: available at https://pdfa.org/resource/glossary-of-accessibility-terminology-in-pdf/

4 Extension schema details

PDF documents using Brotli compression **SHOULD** include in their document catalog dictionary (see ISO 32000, Clause 7.7.2 Document catalog dictionary) an extensions dictionary (in accordance with ISO 32000, Clause 7.12 Extensions dictionary) with a prefix name of **PDFa**. If present, this **SHOULD** contain a developer extensions dictionary in accordance with ISO 32000, Clause 7.12.3 Developer extensions dictionary, with the following entries as shown in Table 1, Developer extensions dictionary values for Brotli.

TABLE 1: Developer extensions dictionary values for Brotli

Key	Type	Value
Type	name	*DeveloperExtensions*
BaseVersion	name	*2.0*
ExtensionLevel	integer	1
ExtensionRevision	text string	2026
URL	string	https://pdfa.org/resource/extension-brotli

NOTE Although effective in reducing file size, the use of Brotli compression on cross-reference streams (see ISO 32000, Clause 7.5.8 Cross–reference streams) or an object stream (see ISO 32000, Clause 7.5.7 Object streams) containing the document catalog may impede processors in locating developer extensions dictionaries.

5 Brotli extension

5.1 General

The PDF specification (ISO 32000) includes an existing filter, **FlateDecode**, that allows for compression of data encoded using the zlib/deflate compression method according to IETF RFC 1951. Brotli is based upon the same LZ77 algorithm that underlies the existing zlib/deflate compression method, thus its use in PDF directly parallels this existing filter.

This document introduces a new filter **BrotliDecode** representing stream data that is compressed using Brotli as defined in IETF RFC 7932. No additional headers or encodings are used.

> **NOTE** The Brotli reference implementation (at GitHub Brotli repository) allows for compression levels from 0 to 11. There is no need to signal the compression level or other options used, as the bitstream is entirely self-contained.

5.2 **BrotliDecode** filter

Append a new row to ISO 32000, Clause 7.4.1 Filters, Table 6—Standard Filters as shown in Table 2, Additions to “Table 6—Standard Filters”:

TABLE 2: Additions to “Table 6—Standard Filters”

Filter name	Parameters	Value
BrotliDecode	yes	(*PDF 2.0*) Decompresses data encoded using the Brotli compression method, reproducing the original text or binary data.

Modify the title of sub-clause ISO 32000, Clause 7.4.4 LZWDecode and FlateDecode filters to be “**LZWDecode**, **FlateDecode**, and **BrotliDecode** filters”.

Append a third bullet and new NOTE to the bullet list in sub-clause ISO 32000, Clause 7.4.4.1 General as follows:

- *The Brotli method is based on the public-domain Brotli compression, in accordance with IETF RFC 7932. Large window Brotli (in accordance with IETF RFC 9841) **SHALL** be supported. No additional signalling is required as the bitstream is self-contained. The use of shared dictionaries (both LZ77-prefix dictionaries, and custom static word/transform dictionaries) are specifically not supported in PDF. The framing format (as defined in IETF RFC 9841) **SHALL NOT** be used.*

> ***NOTE** The standard Brotli static dictionary is used for all streams.*

Modify the title of sub-clause ISO 32000, Clause 7.4.4.3 LZWDecode and FlateDecode parameters to be "**LZWDecode**, **FlateDecode**, and **BrotliDecode** parameters".

Modify the first sentence of the first paragraph in ISO 32000, Clause 7.4.4.3 LZWDecode and FlateDecode parameters to be as follows:

> *The **LZWDecode**, **FlateDecode**, and **BrotliDecode** filters **SHALL** accept optional parameters to control the decoding process.*

Modify the second paragraph in ISO 32000, Clause 7.4.4.3 to be as follows:

> *ISO 32000, Clause 7.4.4.3, Table 8—Optional parameters for LZWDecode, FlateDecode, and BrotliDecode filters shows the parameters that may optionally be specified for **LZWDecode**, **FlateDecode**, and **BrotliDecode** filters. Except where otherwise noted, all values supplied to the decoding filter for any optional parameters **SHALL** match those used when the data were encoded.*

Modify the caption of table ISO 32000, Clause 7.4.4.3, Table 8—Optional parameters for LZWDecode and FlateDecode filters to be "Optional parameters for **LZWDecode**, **FlateDecode**, and **BrotliDecode** filters".

Modify the title of sub-clause ISO 32000, Clause 7.4.4.4 LZWDecode and FlateDecode predictor functions to be "**LZWDecode**, **FlateDecode**, and **BrotliDecode** predictor functions".

BrotliDecode **SHALL NOT** be used for inline images. Modify the paragraph above ISO 32000, Clause 8.9.7 Inline images, Table 91—Entries in an inline image object as follows:

> *These abbreviations are valid only in inline images; they **SHALL NOT** be used in image XObjects. **JBIG2Decode**, **Crypt**, **JPXDecode**, and **BrotliDecode** are not listed in ISO 32000, Clause 8.9.7 Inline images, Table 92—Additional abbreviations in an inline image object because those filters **SHALL NOT** be used with inline images.*

Other than for inline images, **BrotliDecode** may be used where ever **FlateDecode** is permitted.

Bibliography

[1] IETF RFC 1951, DEUTSCH, P. *DEFLATE Compressed Data Format Specification version 1.3*. 1996. RFC Publisher. https://www.rfc-editor.org/info/rfc1951.

[2] GitHub Brotli repository, Google LLC. https://github.com/google/brotli

