The l3pdffile module
Embedding and referencing files in a PDF
LaTeX PDF management testphase bundle

The LaTeX Project*

Version 0.95a, released 2021-02-22

1 l3pdffile documentation

1.1 Introduction

1.1.1 Background

External files can be referenced from a PDF in three ways:

1. through an annotation of type Link,

2. by referencing a local file in the file system,

3. by embedding the file directly into the PDF

Case 1 (Links) are created with the `\pdfannot` commands. This module handles the two other cases. Actually from the view of the PDF format they are quite similar: Case 2 is case 3 without the stream object and without the /EF entry in the /FileSpec dictionary (this points to the stream object of the file). Not embedding the file makes the PDF smaller. But it is also less portable: the files can only be found if they are in the right location relative to the PDF. The normal case is to embed the file.

The tasks to embed and reference such a file are

1. Embed the file in a stream.

2. Create a FileSpec dictionary which references the stream object in the /EF dictionary:

```
<<
 /Type /Filespec
 /F  (l3pdffile.dtx)
 /UF (l3pdffile.dtx)
 /AFRelationship /Source
 /EF <</F 21 0 R /UF 21 0 R>>   %case 3, embedded file
>>
```

The file names in the /UF and /F value don't need to be identical to the name of file on the disc. It is quite possible to embed a `zzz.tex` and name it `blub.tex`. The second name is then what the user will see in the attachment list or in the properties of an annotation.

3. Reference the FileSpec dictionary so that the user can access the file. This can be done in various way:

   (a) With an annotation (/Subtype/FileAttachment). This is done by attachfile, attachfile2 and intopdf. Typical entries of such an annotation are:

key	value type	notes
/FS	object reference	(FileSpec dictionary)
/Name	name	/Graph, /PushPin, /Paperclip, /Tag
/Contents	text string	optional but recommended
/F	integer	Flags
/AP	dictionary	Appearance (required if rectangle >0)
/AS	name	

   The `/AP` takes precedence over Border and similar keys.

   (b) Through an entry in the `/EmbeddedFiles` name tree. This is what embedfiles does.

   ```
   20 0 obj %Document Name tree
    <</EmbeddedFiles 21 0 R>>
        endobj
     21 0 obj %Embedded Files Name dictionary
   <</Names [(AcmeCustomCrypto Protected PDF.pdf) 17 0 R]>>
   endobj
   ```

   The strings (keys) in the `/Names` dictionary must be sorted lexically. But they don't have to be the file name or anything related to the file name. The resource management code uses l3emb0001, l3emb0002 . . . , which allows up to 9999 files. The key can be needed to identify the start file in a collection, so their relation to the files are stored in a property list.

   (c) Through the `/AF` key in various objects (pdf 2.0). The value is normally an array of object references, but it can also be a name which is mapped to an array in /Properties:

   ```
   /AF /NamedAF BDC
   /Properties <</NamedAF [12 0 R]
   ```

   The related `/FileSpec` dictionary should contain an `/AFRelationship` key in this case (but it doesn't harm to add it by default anyway). The values of this key is describe in table 1.

1.1.2 Task 1: Embedding a file

Embedding an existing file is in most cases quite straightforward. This module offers commands, but it can also be done with the basic commands from the `l3pdf` module `\pdf_object_unnamed_write:nn` or `\pdf_object_new:nn`/`\pdf_object_write:nn` or primitive commands to create objects. The object number should be stored for the reference in the `/FileSpec` dictionary.

Table 1: Values of the `/AFRelationship` key

Source	shall be used if this file specification is the original source material for the associated content.
Data	shall be used if this file specification represents information used to derive a visual presentation – such as for a table or a graph.
Alternative	shall be used if this file specification is an alternative representation of content, for example audio.
Supplement	shall be used if this file specification represents a supplemental representation of the original source or data that may be more easily consumable (e.g., A MathML version of an equation).
EncryptedPayload	shall be used if this file specification is an encrypted payload document that should be displayed to the user if the PDF processor has the cryptographic filter needed to decrypt the document.
FormData	shall be used if this file specification is the data associated with the AcroForm (see 12.7.3, "Interactive form dictionary") of this PDF.
Schema	shall be used if this file specification is a schema definition for the associated object (e.g. an XML schema associated with a metadata stream).
Unspecified	(default value) shall be used when the relationship is not known or cannot be described using one of the other values.
Other names	Second-class names (see Annex E, "(normative) PDF Name Registry") should be used to represent other types of relationships.

```
\pdf_object_unnamed_write:nx {fstream}
  {
    {
      /Type /EmbeddedFile
      /Subtype /application\c_hash_str2Fpostscript
      /Params
        <<
          /ModDate  ~ (\file_timestamp:n{example-image.eps})
          /Size     ~  \file_size:n {example-image.eps}
          /CheckSum ~ (\file_mdfive_hash:n {example-image.eps})
        >>
    }
    {example-image.eps}
  }
\tl_set:Nx \l_my_fileobj_tl {\pdf_object_ref_last:}
```

- The `/Params` dictionary is not always required, but the commands of these module will prefill them as shown in the examples. A `/CreationDate` entry has to be added explicitly as there is no sensible way to retrieve this automatically.

- The mimetype (in the `/Subtype`) should be properly escaped. This module contains a property list with maps a number of file extensions to mimetypes and the commands try to detect and fill the mimetype automatically.

- The dictionary can contain additional keys (`/Filter`, `/DecodeParms`), see the pdf reference.

1.1.3 Task 2: Creating the /FileSpec dictionary

The `/FileSpec` dictionary is a simple dictionary object, and can also be created in various ways. If it refers to an embedded file it should reference it in the `/EF` key.

1.1.4 Task 3: Referencing the /FileSpec dictionary

Using the dictionary reference in annotations and `/AF` keys is unproblematic. But to add it to the `/EmbeddedFiles` name tree so that it appears in the attachment panel requires special care: This name tree is a global resource and uncoordinated access can lead to clashes and files that are not visible or inaccessible. The access here is managed by the l3pdfmanagement module:

`\pdfmanagement_add:nnx{Catalog/Names}{EmbeddedFiles}{⟨objref⟩}`

1.2 Commands and tools of these module

file
file/Params
file/streamParams
file/FileSpec

The module predefines and uses a number of local dictionaries for the components of the stream and the `/FileSpec` object. These dictionaries are then used by the `\pdffile_-embed_XX`. The content of these dictionaries can be changed by users with the commands from the l3pdfdict module, but it should be done only locally to avoid side effects on uses by other packages/commands.

The preset values are of these dictionaries are shown in table 2.

Table 2: Preset values in the file dictionaries

dictionary	key	value
l_pdffile	Type	/EmbeddedFile
l_pdffile/Params	Size	\file_size:n{\l_pdffile_source_name_str}
l_pdffile/Params	ModDate	(\file_timestamp:n {\l_pdffile_source_name_str})
l_pdffile/Params	CheckSum	(\file_mdfive_hash:n{\l_pdffile_source_name_str}
l_pdffile/streamParams		a /ModDate entry with year/month/date (used with \pd
l_pdffile/FileSpec	Type	/FileSpec
l_pdffile/FileSpec	AFRelationship	Unspecified

\pdffile_embed_file:nnn \pdffile_embed_file:nnn {⟨*source filename*⟩} {⟨*target filename*⟩} {⟨*object name* ⟩}

This commands embeds the file ⟨*source filename*⟩ in the PDF, and creates a /FileSpec dictionary object named ⟨*object name*⟩. The object name must be unique. The command uses the content of the local dictionaries l_pdffile, l_pdffile/Params and l_pdffile/FileSpec to setup the dictionary entries of the stream object and the /FileSpec dictionary. The /F and /UF entry are filled with ⟨*target filename*⟩.

It is an error if both ⟨*target filename*⟩ and ⟨*source filename*⟩ are empty.

If ⟨*target filename*⟩ is empty ⟨*source filename*⟩ is used instead.

If ⟨*source filename*⟩ is empty, only a /FileSpec dictionary is created.

If the l_pdffile dictionary doesn't contain a Subtype entry with the mimetype, the command tries to guess it from the file extension of ⟨*source filename*⟩. Unknown file extensions can be added (or known extension be changed) by adding to or changing the value in the property \g_pdffile_mimetypes_prop, see below.

When using dvips and pstopdf the actual embedding is done by pstopdf. pstopdf will embed files only if used with the option -dNOSAFER and will not be able to use files which are found with kpathsea.

⟨*target filename*⟩ doesn't need to be a file name with an extension, but it is recommended as security settings in the pdf viewer can restrict access to known file types.

\pdffile_embed_stream:nnn \pdffile_embed_stream:nnn {⟨*content*⟩} {⟨*target filename*⟩} {⟨*object name* ⟩}

This commands embeds the ⟨*content*⟩ in the PDF in a stream objects and creates a /FileSpec dictionary object named ⟨*object name*⟩. ⟨*content*⟩ is wrapped in a \exp_not:n. The object name must be unique. The command uses the content of the local dictionaries l_pdffile, l_pdffile/streamParams and l_pdffile/FileSpec to setup the dictionary entries of the stream object and the /FileSpec dictionary. The /F and /UF entry are filled with ⟨*target filename*⟩. If ⟨*target filename*⟩ is empty the fix name stream.txt is used instead.

If the l_pdffile dictionary doesn't contain a Subtype entry with the mimetype, the command tries to guess it from the file extension of ⟨*target filename*⟩.

⟨*target filename*⟩ doesn't need to be a file name with an extension, but it is recommended as security settings in the pdf viewer can restrict access to known file types.

The stream should not be too long, at least PS imposes a size limit for strings.

\g_pdffile_mimetypes_prop This property contains a list of extensions and their mimetypes. Values can be added or changed with the standard commands:

> \prop_gput:Nnn \g_pdffile_mimetypes_prop {.abc}{text/plain}

The extension should start with a period, the mimetype should be given as plain text (it will be escaped internally). Extensions with two periods are not supported.

\l_pdffile_source_name_str This variable is set at the begin of \pdffile_embed_file:nnn. It can be (and is) used in the file dictionaries, see table 2 for examples.

\g_pdffile_embed_prop This property holds a list of embedded files. It is used by the following show command. The keys are the object names, the argument holds a key word, the source file name and the target file name.

\pdffile_embed_show: This shows the embedded files with their source and target name.

1.3 Example

```
\group_begin:
%set the relationship:
\pdfdict_put:nnn  {l_pdffile/FileSpec} {AFRelationship}{/Source}
%set the description key. The text must first be converted:
\pdf_string_from_unicode:nnN {utf16/string}
   {this~is~an~odd~description~with~öäü}
   \l_tmpa_str
\pdfdict_put:nnx  {l_pdffile/FileSpec} {Desc}{\l_tmpa_str}
%embeds testinput.txt and calls it grüße.txt
\pdffile_embed_file:nnn {testinput.txt}{grüße.txt}{file:example1}
%reference it in the panel
\pdfmanagement_add:nnx
 {Catalog/Names}
 {EmbeddedFiles}
  {\pdf_object_ref:n{file:example1}}
\group_end:
```

2 l3pdffile implementation

```
1 ⟨*header⟩
2 \ProvidesExplPackage {l3pdffile} {2021-02-22} {0.95a}
3   {embedding and referencing files in PDF---LaTeX PDF management testphase bundle}
4 \RequirePackage{l3pdftools}  %temporarly!!
5 ⟨/header⟩

6 ⟨*package⟩
7 ⟨@@=pdffile⟩
8 \cs_new_protected:Npn \__pdffile_filename_convert_to_print:nN #1 #2
9   {\pdf_string_from_unicode:nnN {utf16/hex}{#1}{#2}}
```

2.1 Messages

```
10 \msg_new:nnn { pdffile } { file-not-found }
11   {
12     File~'\tl_to_str:n{#1}'~not~found
13   }
14
15 \msg_new:nnn { pdffile } { mimetype-missing }
16   {
17     Mime~type~not~set~for~file~'\tl_to_str:n{#1}'
18   }
19
20 \msg_new:nnn { pdffile } { target-name-missing }
21   {
22     a~target~name~for~the~/FileSpec~dictionary~is~missing.
23   }
24
25 \msg_new:nnn { pdffile } { object-exists }
26   {
27     object~name~'#1'~is~already~used.
28   }
29
30 \msg_new:nnn  { pdffile } { show-files }
31   {
32     The~following~files~have~been~embedded\\
33      #1
34   }
```

`\l__pdffile_tmpa_tl`
`\l__pdffile_tmpb_tl`
`\l__pdffile_tmpa_str`
`\l__pdffile_tmpb_str`
`\l__pdffile_ext_str`
`\l__pdffile_automimetype_tl`
`\l__pdffile_embed_ref_tl`

temporary variables: generic, for extension, subtype, to store the ref.

(*End definition for* `\l__pdffile_tmpa_tl` *and others.*)

```
35 \tl_new:N  \l__pdffile_tmpa_tl
36 \tl_new:N  \l__pdffile_tmpb_tl
37 \str_new:N \l__pdffile_tmpa_str
38 \str_new:N \l__pdffile_tmpb_str
39 \str_new:N \l__pdffile_ext_str
40 \tl_new:N  \l__pdffile_automimetype_tl
41 \tl_new:N  \l__pdffile_embed_ref_tl
```

`\g_pdffile_mimetypes_prop` This variable holds common mimetypes. The key is an extension with (one) period, the value the description, e.g. `text/csv`.

(*End definition for* `\g_pdffile_mimetypes_prop`. *This variable is documented on page 6.*)

```
42 \prop_new:N \g_pdffile_mimetypes_prop
43 \prop_set_from_keyval:Nn \g_pdffile_mimetypes_prop
44   {
45     ,.csv = text/csv
46     ,.html= text/html
47     ,.dtx = text/plain %or application/x-tex, not in iana.org list
48     ,.eps = application/postscript
49     ,.jpg = image/jpeg
50     ,.mp4 = video/mp4
51     ,.pdf = application/pdf
52     ,.png = image/png
53     ,.tex = text/plain %or application/x-tex, not in iana.org list
```

```
54       ,.txt = text/plain
55       ,.sty = text/plain
56    }
```

`\l_pdffile_source_name_str` \l_pdffile_source_name_str will be set at the begin of the command and contains the full file name and can be used e.g. with \file_timestamp:n.

(*End definition for* \l_pdffile_source_name_str. *This variable is documented on page 6.*)

```
57 \str_new:N  \l_pdffile_source_name_str
```

Here we define and setup the local dictionaries. We add a ModDate to ensure that there is an entry if associated files are used.

```
58 \pdfdict_new:n { l_pdffile }
59 \pdfdict_put:nnn { l_pdffile }{Type}{/EmbeddedFile}
60 \pdfdict_new:n { l_pdffile/Params }
61 \pdfdict_put:nnn { l_pdffile/Params }
62   {ModDate}  { (\file_timestamp:n { \l_pdffile_source_name_str }) }
63 \pdfdict_put:nnn { l_pdffile/Params }
64   {Size}     { \file_size:n { \l_pdffile_source_name_str } }
65 \pdfdict_put:nnn { l_pdffile/Params }
66   {CheckSum} { (\file_mdfive_hash:n { \l_pdffile_source_name_str }) }
67 \pdfdict_new:n { l_pdffile/streamParams }
68 \pdfdict_put:nnn { l_pdffile/streamParams }
69   {ModDate}  {
70                 (
71                   D:\int_use:N\c_sys_year_int
72                    \int_compare:nNnT{\c_sys_month_int}<{10}{0}
73                    \int_use:N\c_sys_month_int
74                    \int_compare:nNnT{\c_sys_day_int}<{10}{0}
75                    \int_use:N\c_sys_day_int
76                 )
77              }
78 \pdfdict_new:n { l_pdffile/FileSpec }
79 \pdfdict_put:nnn { l_pdffile/FileSpec }
80   {Type} { /FileSpec }
81 \pdfdict_put:nnn { l_pdffile/FileSpec }
82   {AFRelationship} { /Unspecified }
83
```

`\g_pdffile_embed_prop` we record here the relation
⟨*object name*⟩ ⇒ {⟨*file/stream or empty*⟩}{⟨*sourcename*⟩}{⟨*targetname*⟩}

```
84 \prop_new:N \g_pdffile_embed_prop
```

(*End definition for* \g_pdffile_embed_prop. *This variable is documented on page 6.*)

`\pdffile_embed_show:`

```
85 \cs_new_protected:Npn \pdffile_embed_show:
86   {
87     \msg_show:nnx
88       {pdffile}{show-files}
89       {
90         \prop_map_function:NN {\g_pdffile_embed_prop} \msg_show_item:nn
91       }
92   }
```

(End definition for `\pdffile_embed_show:`. *This function is documented on page 6.*)

`\pdffile_embed_file:nnn`
`\pdffile_embed_stream:nnn`

At first a command to set the mimetype. It either uses the current value in the file dictionary, or tries to guess it from the extension.

```
93  %#1 file name,
94  %#2 tl to return the (printed) value for the guessed mimetype
95  \cs_new_protected:Npn \__pdffile_mimetype_set:nN #1 #2
96    {
97      \file_parse_full_name:nNNN
98          {#1}
99          \l__pdffile_tmpa_str %unused
100         \l__pdffile_tmpb_str %unused
101         \l__pdffile_ext_str
102       %check if Subtype has been set
103       \pdfdict_get:nnN { l_pdffile}{Subtype}\l__pdffile_tmpa_tl
104       %if not look up in the prop:
105       \quark_if_no_value:NT \l__pdffile_tmpa_tl
106         {
107           \prop_get:NVNTF
108             \g_pdffile_mimetypes_prop
109             \l__pdffile_ext_str
110             \l__pdffile_tmpb_tl
111             {
112               \tl_set:Nx #2 {/Subtype~\pdf_name_from_unicode_e:V \l__pdffile_tmpb_tl}
113             }
114             {
115               \msg_warning:nnx { pdffile }{ mimetype-missing} {#1}
116               \tl_clear:N #2
117             }
118         }
119     }
120
121 \cs_generate_variant:Nn \__pdffile_mimetype_set:nN {VN}
122
123 %#1 file name,
124 %#2 tl, should be empty or contain /Subtype /mimetype
125 %   e.g. result from \__pdffile_mimetype_set:NN
126 \cs_new_protected:Npn \__pdffile_fstream_write:nN #1 #2
127   {
128     \pdf_object_unnamed_write:nx { fstream }
129       {
130         {
131           #2
132           \pdfdict_use:n { l_pdffile}
133           \pdfdict_if_empty:nF { l_pdffile/Params}
134             {
135               /Params
136                 <<
137                   \pdfdict_use:n { l_pdffile/Params}
138                 >>
139             }
140         }
```

`\__pdffile_mimetype_set:nN`
`\__pdffile_mimetype_set:VN`
`\__pdffile_fstream_write:nN`
`\__pdffile_fstream_write:VN`
`\__pdffile_stream_write:nN`
`\__pdffile_stream_write:VN`

9

```
141       { #1 }
142     }
143   \tl_clear:N \l__pdffile_automimetype_tl
144   }
145
146 \cs_generate_variant:Nn \__pdffile_fstream_write:nN {VN}
147
148 %#1 file content
149 %#2 tl, should be empty or contain /Subtype /mimtype
150 %   e.g. result from \__pdffile_mimetype_set:NN
151 \cs_new_protected:Npn \__pdffile_stream_write:nN #1 #2
152   {
153     \pdf_object_unnamed_write:nx { stream }
154       {
155         {
156           #2
157           \pdfdict_use:n { l_pdffile}
158           \pdfdict_if_empty:nF { l_pdffile/streamParams}
159             {
160               /Params
161                 <<
162                   \pdfdict_use:n { l_pdffile/streamParams}
163                 >>
164             }
165         }
166         { \exp_not:n { #1 } }
167       }
168     \tl_clear:N \l__pdffile_automimetype_tl
169   }
170
171 \cs_generate_variant:Nn \__pdffile_stream_write:nN {VN}
172
173 %#1 symbolic name of dict object
174 %#2 target file name,
175 %#3 object ref of the file stream.
176 \cs_new_protected:Npn \__pdffile_filespec_write:nnn #1 #2 #3
177   {
178     \tl_if_blank:nT { #2 }
179       {
180         \msg_error:nn {pdffile}{target-name-missing}
181       }
182       {
183         \pdf_object_new:nn   { #1 } {dict}
184         \group_begin:
185           \__pdffile_filename_convert_to_print:nN  { #2 } \l__pdffile_tmpa_str
186           \pdfdict_put:nnx {l_pdffile/FileSpec}{F} { \l__pdffile_tmpa_str }
187           \pdfdict_put:nnx {l_pdffile/FileSpec}{UF}{ \l__pdffile_tmpa_str }
188           \pdf_object_write:nx { #1 }
189             {
190               \pdfdict_use:n { l_pdffile/FileSpec}
191               \tl_if_empty:nF { #3 }
192                 {
193                   /EF <</F~#3 /UF~#3>>
194                 }
```

```
195              }
196          \group_end:
197      }
198  }
199

200

201  %#1 {source filename}
202  %#2 {target filename}
203  %#3 { filespec object name } (will internally get a prefix!)
204  \cs_new_protected:Npn \pdffile_embed_file:nnn #1 #2 #3
205    { %                  if #1 empty => only filespec
206      %                  if #2 empty => = #1
207      \pdf_object_if_exist:nTF { #3 }
208        {
209          \msg_error:nnn { pdffile }{ object-exists } { #3 }
210        }
211        {
212          \tl_if_blank:nTF { #1 }
213            {
214              \tl_set:Nn \l__pdffile_embed_ref_tl {}
215            }
216            {
217              \file_get_full_name:nNTF {#1} \l_pdffile_source_name_str
218                {
219                  \__pdffile_mimetype_set:VN
220                    \l_pdffile_source_name_str
221                    \l__pdffile_automimetype_tl
222                  \__pdffile_fstream_write:VN
223                    \l_pdffile_source_name_str
224                    \l__pdffile_automimetype_tl
225                  \tl_set:Nx \l__pdffile_embed_ref_tl { \pdf_object_ref_last: }
226                }
227                {
228                  \msg_error:nnn { pdffile }{ file-not-found }{ #1 }
229                }
230
231            }
232          \prop_gput:Nnx
233            \g_pdffile_embed_prop
234            { #3 }
235            {
236              { \tl_if_blank:nTF { #1 } {filespec}{file} }
237              {\l_pdffile_source_name_str}
238              {
239                \tl_if_blank:nTF { #2 }
240                  { \l_pdffile_source_name_str }
241                  { \tl_to_str:n{#2}}
242              }
243            }
244          \tl_if_blank:nTF { #2 }
245            {
246              \exp_args:Nnnx
247                \__pdffile_filespec_write:nnn
248                  %#1 dict, #2 target file name, #3 object ref
```

11

```
249                  { #3 }
250                  { #1 }
251                  {\l__pdffile_embed_ref_tl}
252            }
253            {
254              \exp_args:Nnnx
255                \__pdffile_filespec_write:nnn
256                %#1 dict, #2 target file name, #3 object ref
257                { #3 }
258                { #2 }
259                {\l__pdffile_embed_ref_tl}
260            }
261        }
262    }
263
264
265  %#1{stream content}
266  %#2{target filename}
267  %#3{file object name }
268  \cs_new_protected:Npn \pdffile_embed_stream:nnn #1 #2 #3
269    {
270      %                  if #2 empty => error
271      \pdf_object_if_exist:nTF { #3 }
272        {
273            \msg_error:nnn { pdffile }{ object-exists } { #3 }
274        }
275        {
276            \prop_gput:Nnx
277              \g_pdffile_embed_prop
278              { #3 }
279              {{stream}{}{\tl_if_blank:nTF {#2}{stream.txt}{\exp_not:n{#2}}}}
280          \tl_if_blank:nTF {#2}
281            { \__pdffile_mimetype_set:nN {stream.txt}\l__pdffile_automimetype_tl}
282            { \__pdffile_mimetype_set:nN { #2 } \l__pdffile_automimetype_tl }
283          \__pdffile_stream_write:nN
284            { #1 }
285            \l__pdffile_automimetype_tl
286          \tl_set:Nx \l__pdffile_embed_ref_tl { \pdf_object_ref_last: }
287          \exp_args:Nnxx
288            \__pdffile_filespec_write:nnn
289              %#1 dict, #2 target file name, #3 object ref
290              { #3 }
291              { \tl_if_blank:nTF {#2}{stream.txt}{\exp_not:n{#2}} }
292              {\l__pdffile_embed_ref_tl}
293        }
294    }
295
296
```

(End definition for `\pdffile_embed_file:nnn` *and others. These functions are documented on page 5.)*

```
297  ⟨/package⟩
```

Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.